***CWRU* Action Form for Majors/Minors/Programs/Sequences/Degrees** **Docket #** 14-CSE-PAF-1090

(instructions on back)

College/School:____Case School of Engineering_____
Department: __ Members of the Data Science and Analytics Faculty_____

PROPOSED: **_____ major**
__X___ minor
_____ program
_____ sequence
_____ degree

TITLE: _____Applied Data Science _____
EFFECTIVE: ___Fall_____ (semester) _____2014____ (year)
DESCRIPTION:
Background:

This undergraduate Minor in Applied Data Science (ADS) is based in the Case School of Engineering, and available as a minor to students across CWRU. The ADS Minor will be an important complement to the DSA Major being developed in the Electrical Engineering and Computer Science department of CSE.

The minor is directed to students studying in the domains of Engineering and Physical Sciences (including Energy and Manufacturing), Health (including Translational and Clinical), and Business (including Finance, Marketing, and Economics). Successful completion of the ADS minor requirements leads to a "Minor in Data Science and Analytics" for the graduating student.

The ADS minor represents that the students have developed knowledge of the essential elements of Data Science and Analytics in the area of their major (their domain of expertise). A year long, cross university CWRU research project on data science needs in industry and society showed many educational opportunities for CWRU in this area. The research, data, and findings from this effort are in the supplementary documents in the appendix and are discussed in the Minor Justification.

Elements of the DSA minor:

The minor is structured so that the students who qualify for the minor have a working understanding of the basic ADS tools and their application in their domain area. This includes: 1. Data Management: datastores, sources, streams; 2. Distributed Computing: local computer, distributed computing such as hadoop or other cloud computing; 3. Informatics, Ontology, Query: including search, data assembly, annotation; and 4. Statistical Analytics: tools such as R statistics and high level scripting languages (such as Python).

Students will develop comprehensive experience in the steps of data analysis. Step 1: define the ADS questions, and Step 2: identify, locate, and/or generate the necessary data, including defining the ideal data set and variables of interest, determining and obtaining accessible data and cleaning the data in preparation for analysis. These are followed by Steps 3 through 6. Step 3: exploratory data analysis to start identifying the significant characteristics of the data and information it contains. Step 4: statistical modeling and prediction, including interpretation of results, challenging results, and developing insights and actions. Step 5: synthesizing the results in the context of the domain and the initial questions, and writing this up. And finally Step 6: the creation of reproducible research, including code, datasets, documentation and reports, which are easily transferable and verifiable.

The curriculum is based on five, 3 credit, courses, progressing from Level 1 to Level 5, which cover the spectrum of learning needed to achieve domain area expertise in data science and analytics. The courses are chosen to be both cross cutting, *i.e.*, intermixing students from across the university in the fundamental ADS concepts such as scripting and statistics (Levels 1,2, and 4), and domain area focused (Levels 3 and 5). In addition, Level 4 allows for UG research experience of data science and analytics, in the students' domain area. This meets the 15 credit requirement of a minor.

The ADS minor is a single minor that can be attained by students whose majors are in many different

departments and schools. Their major helps to define their domain area of expertise, and their ADS minor represents their familiarity and learning of the methods of Data Science and Analytics applied in their domain area. The current domain foci can be organized into Engineering and Physical Sciences (*e.g.*, with domain of Astronomy), Energy and Manufacturing, Health (*e.g.*, Translational or Clinical), and Business (*e.g.*, Finance or Marketing).

The data types found in these domains are diverse. They include time series and spectral data for Energy and Astronomy, and sensor and production data and image and volumetric data for Manufacturing. In Health, Translational DSA includes Genomic, Proteomic and other Omics data, while Clinical DSA includes patient data, medical data, physiological time series, and mobile data. Business data types include financial and economic data for Finance, and operations and consumer behavior data for Marketing.

Justification:

See attached Justification, and associated appendices.

This Minor in Applied Data Science is part of the CWRU/BHEF initiative in Undergraduate Data Science, which includes a Data Science and Analytics Major and this Applied Data Science Minor

Is this major/minor/program/sequence/degree: __X___ new
_____ modification
_____ replacement

If modification or replacement please elaborate:
_____

Does this change in major/minor/program/sequence/degree involve other departments?___X__Yes
_____No

If yes, which departments? ___It involves the Data Science and Analytics Faculty, who are from many departments and many schools, including CSE, CAS, SOM, SON, WSOM _____

Contact person/committee: ____Roger H. French __ Roger H. French

SIGNATURES:

DATE
Department James McGuffin-Cawley Digitally signed by James McGuffin-Cawley DN: cn=James McGuffin-Cawley, o=Case Western Reserve University, ou=Materials Science, email=cawley@case.edu, c=US Date: 2014.04.20 09:48:36 -05'00' **Curriculum Chair(s)/Program Directors:**
_____

DATE:
Department Mark R. De Guire Digitally signed by Mark R. De Guire DN: cn=Mark R. De Guire, o=Case Western Reserve University, ou=Department of Materials Science and Engineering, email=mrd2@case.edu, c=US Date: 2014.04.20 21:53:10 -05'00' **Chair:**
_____
__

College/School Curriculum Committee Chair:
Jing Li (please see attached email for approval)
_____

College/School Dean(s):
Gary Wnek (please see attached email for approval)
_____

UUF Curriculum Committee Chair:
_____

File copy sent to: _____Registrar _____Office of Undergraduate Studies/Graduate Studies
_____Other:
_____

From: Jing Li [mailto:jingli@cwru.edu]
Sent: Wednesday, April 30, 2014 9:28 AM
To: Gary Wnek; Jeffrey Wolcowitz
Cc: Roger French; Kathleen Ballou; Alexis Schilf; Anirban Sen Gupta; Camila Estrada; Jing; Kurt Rhoads; LaShanda Korley; Peter Lagerlof; Rohan Akolkar; Scott Hardy; Dwight Davy
Subject: approval of the ADS minor

Dear all,

Based on the feedback from a majority of our committee members, the committee has approved the minor.

Specific comments about course options at each level will be discussed afterwards.

Kathleen,

Is this email sufficient or do you need my signature on the PAF?

--
Jing Li, Ph.D.
Associate Professor
Computational Biology lab @ Case (PI)
Department of Electrical Engineering and Computer Science
Case Western Reserve University

From: Gary Wnek [mailto:gew5@case.edu]
Sent: Sunday, April 27, 2014 7:38 PM
To: Roger French
Cc: Kathleen Ballou; Jing Li
Subject: Re: Applied Data Sciences Minor Action Form

Kathleen can sign for me as I am in England all week.

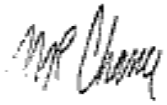
On Apr 27, 2014 11:41 PM, "Roger French" <rxf131@case.edu> wrote:
Hi Jing,

> Hows headway on signing off on the ADS Minor Action Form by the CSE UG Comm.
>
> After that it goes to Gary Wnek, and then its to Jeff Wolcowitz.

Approvals by Department Chairs, whose courses are part of the DSA Minor

Jim McGuffin Signature on Minor Action Form Page
EMSE: Materials Science and Engineering, CSE: Jim McGuffin-Cawley

Hi Roger: It is OK to list ENGR 131, EECS 132 and the new course EECS 133 (which requires approval) in the list of courses for the Data Science minor. Ken
EECS: Electrical Engineering and Computer Science, CSE: Ken Loparo



SYBB: Systems Biology, SOM: Mark Chance

MAMS: Mathematics, CAS: Daniela Calvetti

ASTRO: Astronomy, CAS: Chris Mihos

Justification for Minor in Applied Data Science (ADS)

CWRU and the Business Higher Education Forum initiated a Data Science Workforce Project in July 2013 under the leadership of President Barbara Snyder and Provost Bud Baeslack (Appendix I).

The need for education in Data Science in higher education has been established by this project. Companies of all sizes are using data science and analytics. From the BHEF 2014 Winter Data Science Overview:

> "A 2011 report by McKinsey Global Institute, *Big Data: The Next Frontier for Innovation, Competition, and Productivity*, noted that "big data" is growing at a rate of 40 percent each year and has the potential to add $300B of value to the nation's health care industry alone, with broad application in virtually every sector, as well as scientific organizations and cultural institutions. Projections by Gartner, Inc., indicate that in less than 12 months, 4.4 million IT jobs to support big data will be created globally. About 1.9 million of those jobs will be within the United States, ***and big data has the potential to create three times that number of jobs outside of IT.***" [Emphasis added.]

This project did research on the needs for Data Scientists and the activities at CWRU and peer institutions (Appendix II). In November 2013, an industry roundtable workshop was held on campus to address the needs and demand for Data Scientists and what skill sets and areas of expertise were considered important. (Appendix III).

From this work, Case Western Reserve University and the Case School of Engineering committed to launch Data Science and Analytics Programs. (Appendix IV). In collaboration with corporate partners, Case Western Reserve University is working to launch a degree program in Data Science. The major is to be housed in EECS in the Case School of Engineering and will include a core curriculum of statistics, database systems, data visualization, data privacy, and other data analytics.

The BHEF Data Science Overview (Appendix V) goes on to say that "(w)hile there is a considerable need nationally for data scientists, ***there is an even greater workforce need for the analytics-enabled professional who can marry a deep background in a particular field (e.g., engineering, economics, or business) with a strong understanding of the applications of analytics and visualization tools.***"

The ADS minor will fulfill this need for students with Data Science capabilities who intend to major in fields other than Computer Science, such as Materials Science, Nursing, Management, or Arts and Sciences. The students who graduate with the ADS Minor, will be domain experts with a foundation in the application of Data Science in the domain of their major. Courses will

be available across the university with initial offerings in WSOM, FPBSON, and CAS, in addition to CSE.

The Curriculum: Minor in Applied Data Science

The minor is structured so that the students who qualify for the minor have a working understanding of the basic ADS tools and their application in their domain area. This includes: 1. Data Management: *i.e*., datastores, sources, streams; 2. Distributed Computing: local computer, distributed computing such as hadoop or other cloud computing; 3. Informatics, Ontology, Query: including search, data assembly, annotation; and 4. Statistical Analytics, including tools such as R statistics and high level scripting languages (such as Python).

Students will develop comprehensive experience in the steps of data analysis. Step 1: define the ADS questions, and Step 2: identify, locate, and/or generate the necessary data, including defining the ideal data set and variables of interest, determining and obtaining accessible data and cleaning the data in preparation for analysis. These are followed by Steps 3 through 6. Step 3: exploratory data analysis to start identifying the significant characteristics of the data and information it contains. Step 4: statistical modeling and prediction, including interpretation of results, challenging results, and developing insights and actions. Step 5: synthesizing the results in the context of the domain and the initial questions, and writing this up. And finally Step 6: the creation of reproducible research, including code, datasets, documentation and reports, which are easily transferable and verifiable.

Minor Requirements

The curriculum is based on 5 courses, progressing from Level 1 to Level 5, which cover the spectrum of learning needed to achieve domain area expertise in data science and analytics. The courses are chosen to be both cross cutting, i.e., intermixing students from across the university in the fundamental ADS concepts such as scripting and statistics (Levels 1,2, and 4) and domain area focused (Levels 3 and 5). In addition, Level 4 allows for UG research experience in the students' domain area.

The ADS Minor consists of three cross cutting core courses at Levels 1, 2, and 4 and two domain focused courses (3, 5). Each of these courses is 3 credits, to meet the 15 credit hour requirement of a minor. For the level 4 UG research course, the research topic will be approved by the ADS minor advisor, and will also be a 3 credit project. This will provide ADS minor students both the domain focused ADS learning they need, and a broadening perspective on applications, methods, and uses of ADS in other domain areas.

Level	Physical Sciences & Engineering			Health		Business		
5.	**Physical Sciences** **ASTR 306:** Astronomical Techniques	**Energy** **EMSE 354*** Data Science Modeling and Prognostics for Energy	**Manufacturing** **EMSE 367*** Modeling and Prognostic Data Science for Manufacturing	**Translational** **SYBB 459** Bioinformatics for Systems Biology	**Clinical** **SYBB 322**** Clinical Informatics at the Bedside and Bench Part 2	**Finance**	**Marketing** **MKMR 308** Measuring Marketing Performance **MKMR310** Marketing Analytics	**Economics**
4.	**EMSE 325:** Undergraduate Research **SYBB/BIOL 388:** Undergraduate Research in Biology **ASTR 369:** Undergraduate Research in Astronomy							
3.		**EMSE 353*** Data Science for Energy: Data Sources, Assembly, and Exploratory Data Analysis	**EMSE 366*** Exploratory Data Science for Manufacturing: Data Sources, Assembly, and Exploratory Data Analysis	**SYBB 311**** **A-D** A- Technologies in Bioinformatics B- Data Integration in Bioinformatics C- Translational Bioinformatics D- Programming in Bioinformatics	**SYBB 321**** Clinical Informatics at the Bedside and Bench Part 1		**MKMR 201** Marketing Management	
2.	**STAT 201:** Basic Statistics for Social and Life Sciences **STAT 312:** Basic Statistics for Engineering and Science **SYBB 210**:** Statistics and Tools for Data Science							
1.	**ENGR 131** Elementary Computer Programming **EECS 132:** Introduction to Programming in Java **EECS 133*:** Programming for Data Science							

* CAF in preparation, ** CAF submitted for approval (4xx level already exists)

Cross-listed Courses Counted Toward ADS Minor Requirements

Established courses ready to be included in the ADS Minor are found in CAS (Mathematics, Astronomy, Philosophy); CSE (Materials Science, Electrical Engineering and Computer Science, Manufacturing), SOM, SON, and WSOM (Business and Finance).

Each of the courses that meet the requirements for the ADS Minor can also be taken by students to meet requirements in Major programs, and therefore serve a dual purpose in our academic offerings. Even though each program, department and school may have its own criteria on whether a given course could be "double counted" towards major and minor requirements.

Level 5:

ASTR 306: Astronomical Techniques
EMSE 354*: Data Science Modeling and Prognostics for Energy
EMSE 367*: Modeling and Prognostic Data Science for Manufacturing
SYBB 459: Bioinformatics for Systems Biology
SYBB 322:** Clinical Informatics at the Bedside and Bench Part 2
MKMR 308: Measuring Marketing Performance
MKMR 310: Marketing Analytics

Level 4: (Subject to approval by DSA minor advisor)

EMSE 325: Undergraduate Research
SYBB/BIOL 388: Undergraduate Research in Biology
ASTR 369: Undergraduate Research in Astronomy

Level 3:

EMSE 353*: Data Science for Energy: Data Sources, Assembly, and Exploratory Data Analysis

EMSE 366*: Exploratory Data Science for Manufacturing
SYBB 311 A-D**, A- Technologies in Bioinformatics, B- Data Integration in Bioinformatics, C- Translational Bioinformatics, D- Programming in Bioinformatics
SYBB 321:** Clinical Informatics at the Bedside and Bench Part 1
MKMR 201: Marketing Management

Level 2:

STAT 201: Basic Statistics for Social and Life Sciences
STAT 312: Basic Statistics for Engineering and Science
SYBB 210:** Statistics and Tools for Data Science

Level 1:

ENGR 131 Elementary Computer Programming
EECS 132: Introduction to Programming in Java
EECS 133*: Programming for Data Science

* CAF in preparation, ** CAF submitted for approval (4xx level already exists)

Domains

The ADS minor is a single minor that can be attained by students whose majors are in many different departments and schools. Their major helps to define their domain area of expertise, and their DSA minor represents their familiarity and learning of the methods of Applied Data Science in their domain area. The current domain foci can be organized into Engineering and Physical Sciences (*e.g.*, with domain of Astronomy), Energy and Manufacturing, Health (*e.g.*, Translational or Clinical), and Business (*e.g.*, Finance or Marketing).

The data types found in these domains are diverse. They include time series and spectral data for Energy and Astronomy, and sensor and production data and image and volumetric data for Manufacturing. In Health, Translational ADS includes Genomic, Proteomic and other Omics data, while Clinical ADS includes patient data, medical data, physiological time series, and mobile data. Business data types include financial and economic for Finance, and operations and consumer behavior data for Marketing.

Applied Data Science Minor Initiating Faculty

The ADS Minor is based in the Case School of Engineering, and is founded by the initial ADS Minor Faculty from schools across the university. The current members of the ADS Minor Faculty are as follows.

	ADS Faculty	
Name	Dept.	School
Roger French	EMSE	CSE
GQ Zhang	EECS	CSE

Jim McGuffin-Cawley	EMSE	CSE
Alexis Abramson	EMAE	CSE
Mark Chance	SysBio	SOM
Colin Drummond	Healthcare IT and Informatics	SON
Jagdip Singh	Design & Innovation	WSOM
Robin Dubin	Economics	WSOM
William Mahnic	Finance	WSOM
Jennifer Carter	EMSE	CSE
Chris Mihos	Astronomy	CAS

Academic Advisement for Minor in Applied Data Science

Existing policy regarding academic advising roles states that academic / departmental representatives who function as faculty advisors for majors and minors have "specific knowledge of the department; connection with faculty within a department, and knowledge of opportunities for engagement within the department; and familiarity with major / minor requirements". Therefore, academic advisement of undergraduate students seeking a minor in ADS will be based in the departments of the faculty teaching courses in the minor.

Applied Data Science

Minor Action Form

Appendices

Appendix I. Case Western Reserve University and Business Higher Education Forum Data Science Workforce Project

Appendix II. Data Science Matrix for Case Western Reserve University Peer Institutions

Appendix III. Data Science Industry Roundtable – Monday, November 4, 2013: Summary

Appendix IV. Project Statement: New Undergraduate Degree Program in Data Science

Appendix V. BHEF Winter 2014 Meeting: Data Science Overview

Appendix VI. Key Word Glossary: Data Science & Analytics

Appendix VII. Syllabi

Case Western Reserve University and Business Higher Education Forum
Data Science Workforce Project

Companies and governments use data science and advanced analytics to drive innovation, shape strategy and inform decision-making.

The field of **data science** addresses this need. Combining mathematics, statistics, data engineering, computing and other disciplines, data science uncovers competitive advantages, new opportunities, customers' needs, or previously unforeseen challenges.

Imagine what data science could mean to your business.

Our faculty has, and Case Western reserve is responding with new undergraduate programs in data science. Numerous studies underscore the need:

- McKinsey Global Institute predicts a **nationwide shortage of 140,000 to 190,000 workers** with "deep analytical skills," and **a deficit of 1.5 million managers** capable of using big data analytics for actionable insights in their decision-making.
- McKinsey predicts a **40 percent annual growth in global data** and $**300 billion in potential value-add** of data analytics to the nation's health care industry alone.
- Gartner Inc. indicate that by 2015, **4.4 million information technology jobs** globally will be created to support data science and analytics, generating **1.9 million IT jobs** in the U.S.,
- The U.S. government anticipates a **shortage of about 50,000 qualified workers** in health IT between 2010 and 2015. Healthcare companies are increasingly borrowing technology specialists from other industries.

Case Western Reserve is collaborating with the Business-Higher Education Forum to create exciting programs in data science, and we need your help. What skills do you need from our graduates? How can we tailor our program to best meet your needs? We seek you input on our curriculum and your support for research internship and co-op opportunities for students at your organization.

Collaboration with corporate partners is a powerful tool. We recently launched three new master's programs in our School of Engineering based on feedback from 60 companies. We aim to replicate that success here, while creating a national model for undergraduate education in data science.

About Case Western Reserve University

Case Western Reserve University is one of the country's leading private research institutions. Located in Cleveland, we offer a unique combination of forward-thinking educational opportunities in an inspiring cultural setting. Our leading-edge faculty engage in teaching and research in a collaborative, hands-on environment. Our nationally recognized programs include arts and sciences, dental medicine, engineering, law, management, medicine, nursing and social work. About 4,200 undergraduate and 5,600 graduate students comprise our student body. Visit case.edu to see how Case Western Reserve thinks beyond the possible.

About the Business-Higher Education Forum

Now in its 35th year, BHEF is the nation's oldest membership organization of Fortune 500 CEOs and research university presidents dedicated to advancing innovative education and workforce solutions and improving U.S. competitiveness. BHEF's business and academic members collaborate in regions across the country to design and deploy education-workforce solutions in the high-demand and emerging fields that are so critical to innovation and national security. BHEF and its members drive change locally, work to influence public policy at the national and state levels, and inspire other leaders to act.

BHEF BUSINESS-HIGHER EDUCATION FORUM®

Creating Solutions.
Inspiring Action.

Data Science Matrix for Case Western Reserve University Peer Institutions

October 30, 2013

Methodology

- Worked with CWRU to develop list of 22 peer institutions
- Collected data on Degree Programs, Departments, Centers, and Courses supporting data science
  - searched for terms “data science,” “data analytics,” “data mining,” and “informatics” for each university
  - searched through relevant program curriculums and courses for each university
  - conducted university-specific searches based on data science research expertise

Peer Matrix is a Work in Progress...

- Many university websites are difficult to navigate
  - Course listings were occasionally password protected or not centrally located for a department
  - Centers/Institutes rarely listed in a central location
- Data science courses scattered throughout University
  - Difficult to find courses with field-specific data science courses (e.g., in social science or astronomy)
- Names of industry partners not usually recognized on University websites
  - May require specific searches to find industry partners

Commonly Offered Courses in Data Science

- To create an data science expert graduate, there are several "core" courses that are offered at multiple institutions
  - data structures
  - database systems
  - data mining
  - statistics
  - computer vision
  - machine learning
- To create a data science enabled graduate, there are a few "core" courses that are offered at multiple institutions
  - business analytics
  - bioinformatics
  - health informatics
  - computational biology
  - genomics

Conclusions on Data Science Programs

- Nearly all data science courses are located within the Computer Science Department
  - Other common choices include Information Science; Computer Science & Engineering; Bioinformatics; School of Business
- Nearly all data science education is targeting the graduate level
  - Undergraduate education targets upper-level students
- Courses and curriculum vary significantly across institutions
  - Data science programs support research interests/strengths of university
- No clear common understanding of a “data science training/education” at the undergraduate or graduate level
  - At graduate level, training/education is highly field specific

Implications for CWRU

- CWRU has the opportunity to be a national leader in defining a "core" curriculum in data science, particularly at the undergraduate level
  - Foundational topics could include statistics, database systems, data visualization, pattern recognition, data privacy, data analytics
- Data science concentrations within current disciplinary areas can create a significant number of "data science enabled" graduates
  - Inclusion of a hands-on learning experience, such as a co-op or internship, demonstrates to employers that graduates have applicable skills
- CWRU can build on its strengths to expand data science education/training and create ties with industry
  - Health sector represents a critical area of comparative advantage for Cleveland and Northeastern Ohio
  - Other sectors in Cleveland also have a strong demands in data science

Higher Education	School/ College/	Centers/Institutes/Programs	Major/ Minor/ Certificate	Special Topics Areas	Courses Included	Cohort Size	Lecture	Funding or Partnerships
Boston University	Computer Science; Biology; Epidemiology (graduate)		Concentrations within Computer Science BS, MS, PhD; Graduate certificate in Database Management and Business Intelligence; Graduate certificate in Health Informatics;	bioinformatics; biostatistics; computational biology; systems biology;	introduction to databases and data mining; introduction to database systems; data mining; pattern matching and pattern detection; machine learning; database applications; biological database systems; statistical pattern recognition; advanced statistical methods; mathematical and statistical methods of bioinformatics; data structures and basic algorithms;		Database and data mining reading group	
Case Western Reserve University	Biology; Biomedical Engineering; Electrical Engineering and Computer Science; Epidemiology and Biostatistics; Genetics;	Cancer Center; Center for Proteomics and Bioinformatics	PhD in Information Systems; MS in Wireless Health; MS in Biology and Bioinformations; PhD in Biology and Bioinformatics	informatics; bioinformations; computational and molecular biology; translational bioinformatics; health				
Carnegie Mellon University	School of Information Systems and Management; School of Public Policy and Management; Department of Biomedical Engineering; Machine Learning Department (used to be Center for Automated	iLab (interdisciplinary research center including Department of Statistics, Department of Machine Learning, School of Computer Science and Tepper School of Business); Living Analytics Research Center; Center for Future of Work; Center for Bioimage Informatics (graduate education and undergraduate research internships); Lane Center for Computational Biology	Master of Information Systems Management (MISM) with concentration in Business Intelligence and Data Analytics; Master of Entertainment Industry Management (MEIM); MS in Health Care Policy and Management; MS in Biotechnology and Management; MS in Medical Management; MS in Information Technology Systems with a concentration in Big Data and Analytics; PhD in Machine Learning; Undergraduate minor in Machine Learning	predictive modeling; GIS mapping; data visualization; social media; health informatics; search engines and web mining; information retrieval; bioinformatics; artificial intelligence; computer vision; language technologies; neural	distributed systems; database management; object oriented programming in Java, data mining, analytics & business intelligence; data warehousing; digital transformation; measuring social; applied econometrics; data analysis in health fields; business process modeling; large scale data analysis; text analytics; MEIM courses include database management; decision analysis; IT management, technology policy; machine learning and policy; statistical theory for social and policy research; event and pattern detection; data analysis for managers; information systems for managers; large scale data analysis for public policy; health care information systems; SAS for policy management; business intelligence and data mining SAS; introduction to Raster GIS; algorithms and data structures for Information processing; data structures; data mining; data warehousing; exploring and visualizing data; multimedia		Database seminar series, CyLab seminar series	Many informal relationships for internships; $560,000 gift of equipment from IBM Corp for CALD
Columbia University	Offered jointly through Graduate School of Arts and Science and School of Engineering and Applied Science	Institute for Data Sciences and Engineering (comprised of Center for New Media, Center for Health Analytics, Center for Financial Analytics, Center for Foundations of Data Science, Center for Smart Cities, Center for Cybersecurity)	Certification of Professional achievement in Data Sciences; MS to begin enrolling fall 2014; PhD fall 2015; not open to currently enrolled Columbia students for credit		algorithms for data science; probability and statistics; machine learning; exploratory data analysis and visualization; data engineering; statistical inference and modeling; data science capstone course; introduction to data science (team taught with Google employee).	First class enrolled fall 2013; hiring 30 new faculty and up to 75 new faculty	Monthly colloquium series	Part of NYC's Applied Sciences NYC initiative; NYC provided $15M in financial assistance
Emory University	Business School (Information Systems and Operations Management or Marketing Concentration); School of Medicine Department of Biomedical Informatics; Math/Computer Science Biomedical Informatics	Marketing Analytics Center; Center for Comprehensive Informatics; Biomedical Informatics Program (collaborative effort of Emory University, Morehouse School of Medicine, and Georgia Institute of Technology focused on linkages and integration of translational and clinical/health systems such as existing and new electronic records systems,	MBA focus area; BBA focus area; PhD and MS in Biomedical Informatics, MPH or PhD in Public Health; PhD in Pathology Informatics	imaging informatics; clinical and translational research informatics; high-end computing for informatics/analytics/big data	data and decision snalytics; consulting simulation analysis; data analytics and visualization; foundations of digital markets; health care operations and technology management; business forecasting and predictive analysis; analytics for e-markets; management science in spreadsheets; marketing analytics; database systems; introduction to biomedical informatics; exascale data analytics; high performance computing; advanced database systems; database mining; biomedical image analysis; introduction to clinical analytics; principles of public health informatics; management principles for informatics; geographic information systems; statistical methods; biostatistical methods; applied linear models; fundamentals of machine learning; machine learning and computational biology, modern regression analysis.			
Georgia Tech	College of Computing (Department of Computational Science and Engineering);	Institute for Data and High Performance Computing; Georgia Tech Research Institute; Foundations on Data Analysis and Visualization Analytics Research	MS or PhD in Computational Science and Engineering; MS or PhD in Electrical and Computer Engineering	machine learning; bioinformatics	data and visual analytics; computational data analysis; high performance computing; web search and text mining; pattern matching; computer visualization techniques; computational statistics; financial data analysis; data mining and statistical learning; algorithms for bioinformatics and computational biology; computability, algorithms, and complexity; high			
Lehigh University	College of Engineering (Computer Science and Engineering)		BS, MS, or PhD in Computer Science and Engineering (concentrations in artificial intelligence, bioinformatics, computing		pattern recognition; data mining; bioinformatics; database systems; biomedical image computing modeling			
Massachusetts Institute of Technology	Computer Science, Electrical Engineering and Computer Science, Sloan School of Management	Computer Science and Artificial Intelligence Laboratory (CSAIL)	BS, MS, PhD Computer Science; BS, MS, PhD in Electrical Engineering and Computer Science; MS, PhD in Biomedical Informatics; MS in Management Science; Minor in Management Science; MBA or MS in Management; MS in Finance; MS in Management and Engineering;	finance; medicine; social media; security	artificial intelligence; robotics; machine vision; human intelligence; data structures; machine learning; medical artificial intelligence; health informatics; processing, analyzing, and visualizing data; data mining; artificial intelligence; machine learning; robotics, inference and information; algorithms for inference; information theory; computational systems biology; human intelligence enterprise; computational cognitive science; database systems; cognitive robotics; distributed algorithms; advances in computer vision; cryptography and cryptanalysis			Intel Science and Technology Center for Big Data at CSAIL; AIG; EMC; Intel; SAP; Thomson Reuters

New York University	Business School; Biology; Mathematics; Computer Science; Physics; Economics; Statistics	Initiative in Data Science and Statistics	MS in Data Science (first cohort started in fall 2013); MS in Applied Urban Science and Informatics; MA in Applied Quantitative Analytics; MBA with a Specialization in Business Analytics; MS in Business Analytics; MS in Applied Statistics for Social Science Research; PhD in Biostatistics; MS in Scientific Computing Data; PhD in Computer Science (Machine learning/Artificial Intelligence Specialization); PhD in Statistics; Advanced Certificate in Applied Urban Science and Informatics; MS of PhD in Information Systems; MS or PhD in Mathematics; MS or PhD in Biology; MA in Economics; MS or PhD in Physics; PhD in Music Technology	developing and utilizing automated methods of analyzing data	data science; statistical and mathematics methods; machine learning and computational statistics; big data; inference and representation			
The Ohio State University	College of Medicine (biomedical Informatics); College of Engineering (Department of Computer Science and Engineering and Department of Electrical and Computer	Discovery Themes Initiative (ten-year, multi-million dollar investment designed to help Ohio State attract new tenured and tenure-track faculty in the Discovery Theme areas, using data analytics as a foundational tool for tackling the challenges within their Theme);	MS, PhD, MPH, in Biomedical Informatics; MS, BS,BA, or minor in Computer and Information Science; BS or MBA (Business Administration with Information Systems, Business Analytics, or Management Sciences option); BS (Electrical and Computer Engineering); PhD in Biostatistics	integration of large-scale data analysis; management, processing, and visualization with biomedical informatics; modeling and problem solving with spreadsheets	introduction to bioinformatics; biomedical informatics; imaging informatics; research informatics; database systems; advanced database management systems, data structures and algorithms; pattern recognition and machine learning, computer vision; regression analysis; multivariate analysis; Bayesian analysis; statistical analysis of genetic data; introduction to genomic methods; analysis of high content and high throughput data for neuroscience; invitation to probability; statistical phylogenetics; business analytics; business statistics; decision sciences; information systems planning and management			
Penn State University	Statistics; Computer Science & Engineering; Information Sciences & Technology; Geography; Political Science; Sociology; Anthropology; Economics;	IGERT Program in Big Data in Social Science; "Hacking Science" Blog to cover coding, big data, and science at Penn State; Institute for Cyber Science (uses computational cyberscience in areas of Energy and the Environment, Life Sciences, Materials, and Social	PhD or PhD minor in Social Data Analytics; MS or PhD minor in Computational Cyberscience; Graduate certificate in Applied Statistics (on-line)	social science	analytics; data mining; machine learning; visual analytics; social data analytics; privacy in statistical databases; big social data and the law; the information environment; data privacy; learning and games; network science; vision-based tracking; computational regularity; large data sets; pattern recognition; web analytics; spatial analysis; information retrieval and organization; statistics; social network analysis; democratic representation: big data approaches; spatial demography; geospatial science in anthropology; biological data analysis; networks in life science; concurrent scientific computing; concurrent matrix			
Rensselaer Polytechnic Institute	Information Technology; School of Management;	Data Science Research Center; Exploratory Center for	BS, MS, or PhD minor in Information Technology and Web Science; MS in Business analytics		information systems for management; data resource management; knowledge discovery with data mining; predictive analytics using social media;technology fundamentals for business	Launched September 2013	monthly seminars	GE Global Research; IBM
Stanford University	Statistics; Computer Science; Biomedical computation	Center for Clinical Informatics	Mining Massive Data Sets graduate certificate (school of professional development; MS in Biomedical Informatics; MS in statistics; BS,MS, or PhD in Computer Science; BS in Biomedical Computation		social and information network analysis; machine learning; mining massive data set; information retrieval and data search; biomedical informatics; modeling biomedical systems; representations and algorithms for computational molecular biology; translational bioinformatics; biomedical image analysis and representation; machine learning; artificial intelligence; data mining; image analysis; human-computer interaction; systems engineering; scientific and numerical computing; data driven medicine; data mining and analysis; computer			
University of California Berkeley	School of Information	Affiliated with Center for Law and Technology; Center for New Media; Blum Center for Developing Economies; Center for Information Technology Research in the Interest of Society; Townsend Center for the Humanities; Electronic Cultural Atlas Initiative; Information and Communication Technologies and Development; D-Lab (helps Berkeley faculty, staff, and graduate students move forward with world-class research in data intensive social science)	MS in Information Management and Systems; MS in Information and Data Science; PhD in Information Management and Systems	information and computer science design; social sciences, management human-computer interaction; information economics and policy; information law and policy; information organization and retrieval; information systems design; social aspects of information; information and communication technologies and development	history of information; social and organizational issues of information; information law and policy; needs and usability assessment; managing in information-intensive companies; information technology economics, strategy, and policy; cyberlaw; information visualization and presentation; technologies for creativity and learning; information systems and health care; designing mobile experiences; data mining and analytics in intelligent business services; alternative visions of technology; health care and strategy; media, new and otherwise; working with open data; information access; data science and analytics: thought leaders; web architecture and information management; technology and poverty; information organization and retrieval; privacy, security, and cryptography; user interface design and development; database management; quantitative research methods for information systems and management; finding health in the US; health care and the information economy; social data revolution	On-line MS in Information and Data Science to enroll first class of 30 students January 2014; Currently more than 100 students enrolled in MS in Information Management and Systems; 20-25 in PhD program	Some courses are seminar series courses	

University of California Los Angeles	Information and Computer Sciences; Information Studies	Host "DataFest" (Big Data competition for undergraduates); Clinical and translational science institute	BS or minor in Informatics; BS in Business Information Management; BS in Information and Computer Science; Minor in Statistics; MS in Library and Information Studies (specialization in	data mining	bioinformatics; network fundamentals; artificial intelligence; computational and systems biology; pattern recognition and machine learning; data and decisions; data mining; database theory; database management systems; financial data mining; information studies; information structures; information retrieval; information architecture; information systems		
University of California San Diego	UCSD School of Medicine (Public Health); School of Management	San Diego Supercomputer Center, Health Services Research Center, Center for Large-scale Data Systems Research,	Data Mining Certificate (online); Biostatistics certificate(online); MBA training (no specific certificate); BS in Computer Science (specialization in bioinformatics); PhD	system architectures and software; analytics; performance; data growth and use of data; data	data mining; data preparation for data mining; hadoop essentials; R programming; predictive analytics; DAD programming; data mining for scientific applications; biostatistics; competitive advantage through cloud computing; business aspects of cloud computing; technical aspects of cloud computing; advanced data structures; database system principles; database system	seminar series	
University of Chicago	School of Business; Computer Science	Host "Data Science for Social Good Summer Fellowship Data Slam" (competition for undergrads and grads); Computation Institute; Conte Center for Computational Neuropsychiatric	MBA Analytical Management concentration; BS, MS, or PhD in Computer Science		managerial decision modeling; data mining; strategic business investments; data driven marketing; database systems; scientific visualization; artificial intelligence; computational linguistics; machine learning and large scale data analysis; computer vision; machine learning; computational biology; database systems; networks and distributed systems; digital biology		
University of Illinois Urbana-Champaign	Computer Science	Multimodal Information Access & Synthesis Center (within Computer Science Department); CyberGIS Center	Data Sciences Summer Institute; BS in Computer Science; MS in Computer Science; MS in Bioinformatics; PhD in Computer Science		structured prediction; text information systems; linguistic structure; computational linguistics; natural language processing; linguistic analysis; speech and language processing; information storage and retrieval; mathematical models of language; information theory;	seminar series	Alcatel-Lucent Bell Labs; AT&T labs; Geosemble Technologies
University of Rochester	Computer Science; School of Medicine and Dentistry; Business School	Institute of Data Science ($100M Commitment); Health Science Center for Computational Innovation; Center for Integrated Research Computing;	BA, BS, MS, PhD in Computer Science; MS in with focus on Data Mining; MS in Business with a focus on Business Analytics (partnership with Engineering and Applied Science); MS in	predictive health analysis; cognitive systems and artificial intelligence; analytics on demand	data structures; science of data structures; explorations in robotics; machines and consciousness; computational models of music; human computer interaction; artificial intelligence; natural language processing; statistical speech and language processing; cryptography; optimization and machine learning; cognitive science; digital media; machine	annual Big Data Forum, Data Analytics symposium,	IBM; Xerox
University of Washington	Statistics; Biostatistics; Applied mathematics	eScience Institute; IGERT Program	Certificate in Data Science (continuing education); Certificate in Cloud Data Management and Analytics (continuing education); Certificate in Business Intelligence: Techniques for Decision Making (continuing education); MOOC in Data Science; PhD in Computer science with specialization in Big Data (new September 2013); BS or MS in Computer Science; plan to have an MS in eScience (data science); Certificate in Statistical Analysis with R programming; PhD in Statistics with specialization in Big data		data science; data analysis; deriving knowledge from data at scale; information technology; data programming; data abstractions; data management; data structure and algorithms; computational biology; database systems internals; machine learning; distributed systems; computer vision; very large scale integration; computational linguistics; artificial intelligence; synthetic biology; data compression; data mining; database management; machine learning; machine learning for big data; probabilistic and statistical databases; data visualization; scalable analytics in the cloud; cloud computing; numerical methods for astrophysics		
University of Wisconsin-Madison	Statistics; Biostatistics; Biometry; Computer Science	Data and Information Services Center; Data Mining Institute (Computer Science); Space Science and Engineering Center	MS or PhD in Biostatistics; MS in Biometry; BS, MS or PhD in Statistics; Bioinformatics Certificate		regression analysis; statistics; data analysis; multivariate analysis; statistical data processing; biostatistics; statistical methods for bioscience; epidemiology; statistical computing; data structures; cryptography; computational statistics; combinatorics; pattern recognition; image processing; artificial neural networks and fuzzy systems; artificial intelligence; natural		Microsoft
Vanderbilt University	Scientific Computing; Astronomy; Biology; Economics; Mathematics; Psychology; School of Medicine	Initiative in Data-Intensive Astrophysics; Center for Quantitative Sciences; Informatics Center	Minor in Scientific Computing (undergraduate); PhD Astronomy; BS, MS or PhD in Computer Science; BS in Computer Science with Financial Engineering focus; MS or PhD in Biomedical Informatics; BS, MS, or PhD in Electrical		program design and data structures; high performance scientific computing; stellar astrophysics; galactic astrophysics; computational genomics; genome science; model living systems for biomedical engineering; molecular modeling methods; econometrics; mathematical modeling in economics; scientific computing for psychological and brain science; biomedical informatics; clinical information systems and databases; systems		
Washington University in St. Louis	Biomedical Engineering; Computer Science and Engineering	Center for Biomedical Informatics; Neuroimaging Informatics and Analysis Center	BS in Biomedical Engineering; Minor in Bioinformatics; MS and PhD in Human and Statistical Genetics; BS, MS, and PhD in		biomedical Informatics; computational statistical genetics; bioinformatics; computational molecular biology; data mining and applications to molecular biology; genomics; statistics; population genetics; algorithms and data structures; social network analysis; artificial	seminar series	

Case Western Reserve University / Business-Higher Education Forum

Data Science Industry Roundtable – Monday, November 4, 2013

Breakout Session Roundup - Overview

Shared Current Hire Characteristics:

- Companies are hiring *both* recent and mid-career graduates from a range of data science-related fields, including: engineering, statistics, software development, business analytics, etc..
- Appreciation for seasoned employees who have experience with current and legacy computer systems and tools.
- Strong desire to have new hires with real-world experience.

Shared Skill Gap Assessments:

- Need for better communication skills (speaking, writing, presenting, and convincing).
- Need for more hands-on experience before day one.
- Need for increased understanding of the "business case" behind data science applications – including finance and economics.

Shared Suggestions:

- New hires should be able to think critically and communicate well: ask questions, explore data, defend positions, work in a team and accept new direction.
- Data scientists don't need to be subject matter experts, but they need to know *some* and learn *more* specific to industry sector.
- Pair undergrads with more experienced employees.
- Offer/create more internships, co-ops, real world experiences.

Case Western Reserve University / Business-Higher Education Forum

Data Science Industry Roundtable – Monday, November 4, 2013

Breakout Session Roundup – 1

Services:

The Services group outlined key skills a data science enabled employee would possess:

Characteristics of Current Hires:

- Mathematicians
- Statisticians
- Engineers

Existing Skills Gap not being addressed:

- Data Science focus with business/discipline minor
- Required understanding the full data lifecycle
- Need 21st century competencies
- Ability to deal with unstructured environments and problems
- Awareness of computing infrastructure
    - current and legacy systems
- Have managerial courage in reporting

Suggestions

- Develop skills beyond only coding
- Encourage structured thought processes
- Train in analytical thinking
- Provide exposure to behavioral sciences
- Offer breadth of real-world experiences
- Encourage understanding of data through industry-specific lens
- Develop both types of people: right and left brain
    - Those who can ask the questions
    - Those who can crunch data

Case Western Reserve University / Business-Higher Education Forum

Data Science Industry Roundtable – Monday, November 4, 2013

Breakout Session Roundup – 2

Health Care:

The Health Care group focused on characteristics of current hires, unaddressed workforce needs, and suggestions for future workforce development strategies.

Characteristics of current hires:

- Graduates of Tier I R&D Universities
- Enabled w/ Data Science Skills
  - Hacking/programming/Big Data experience
  - Visualization (undergrad)
- Hired from other companies (looking for well-rounded/domain expertise)
- Often, seasoned employees
  - Good w/ tools already used in industry
  - Project management experience

Existing Skills Gap Not Being Addressed:

- Need soft skills – Communication
- Need the ability to talk about data-enabled decision-making process
- Need for better presentation skills
- Need "enabled" graduates w/ experience using data tools, not just project management experience

Suggestions:

- Pair undergrads with more experience workforce
- Need an increase in work experience (co-ops, internships)
- Need to focus on real-world applications: Is there a new way to adapt the curriculum to address real-world needs?
- Focus on team-based learning
- Increase industry collaboration and academic partnerships
- Remember: Data scientists don't need to be SME's, but they need to know a little and need to be able to learn/acquire more

Case Western Reserve University / Business-Higher Education Forum

Data Science Industry Roundtable – Monday, November 4, 2013

Breakout Session Roundup – 3

Energy/Manufacturing

The Energy/Manufacturing group focused on characteristics of current hires, the existing skill gap, and the ideal future candidate.

Characteristics of current hires:

- Both mid-career and recent graduates from grad and undergrad (MS/BS) programs
- Background in engineering
  - Computer Science, IT, EE, Power Engineers, Software Engineers
- Background in business/finance

Existing Skills Gap Not Being Addressed:

- Lack understanding of Business Case
- Often lack domain expertise – consumer behavior, integrated operations, business processes
- Need more hands-on application experience
- Increase communication skills – listen, talk, write

Suggestions:

- Create both "expert" and "enabled"
- Build understanding of business side - Economics/Finance
- Need "core" skill set
  - Information security
  - Ethics
  - Broad tool kit
  - Advanced statistics
  - Basing engineering knowledge
  - Communication skills

CASE WESTERN RESERVE UNIVERSITY EST. 1826

Project Statement:

New Undergraduate Degree Program in Data Science

"Leaders in business, education and government must take action to foster a new generation of talent with the technical expertise and unique ideas to make the most of this tsunami of Big Data."

~Richard Rodts, Manager of Global Academic Programs, IBM

The volume of data acquired or generated by corporations continues to increase at unprecedented rates. The need to efficiently interpret data into information useful for strategic decisions is paramount. For example, information within these large data sets can lead to better assessment of customers' needs, business decisions could be made with more confidence, and opportunities for greater operational efficiencies, cost reductions and reduced risk could be identified. All of these aspects have become increasingly important in today's global business community. The potential impact of data analytics on the way industry does business is akin to the way the internet has shifted paradigms for commerce.

While the development of analytics skills for today's professionals has become a necessity, there exists a tremendous gap in the growing number of jobs that require data-driven skills, and the available pool of job candidates who can fulfil these roles. In fact, according to a recent IBM *Tech Trends Survey, "*only one in ten businesses have the required skills to use state-of-the-art technology in the field of Big Data and analytics."

The Program

Case Western Reserve is launching a distinctive undergraduate degree program in data science in the fall of 2014.

Rooted in a core curriculum of data-science fundamentals, our program will prepare graduates to meet industry needs through cutting-edge instruction, exciting areas of concentration and an extensive capstone experience.

Students pursuing this curriculum will receive a strong foundation in topics fundamental to data science (such as mathematics, statistics, and computing algorithms), breadth in allied technical areas (such as biology, physics, or engineering), and a focus on a specific domain area: health, energy, and manufacturing and production. As the program matures, other concentrations such as finance may be added.

All of our data science programs will include a strong experiential learning component. A minimum of two internships in data science (one in the summer between sophomore and junior year, the second in the summer between junior and senior year) or one seven-month, full-time co-op assignment will be expected. In addition, all students must complete a two-semester senior capstone project in data science.

Students completing our data science programs will be well prepared to fill entry-level workforce needs for data scientists in a variety of industries. We also expect to launch a post-baccalaureate certificate program for industry personnel interested in retraining.

Corporate partners

Case Western Reserve is forming partnerships with businesses to:

- Help build curriculum in data science and analytics
- Provide new infrastructure to support a pipeline of students – both undergraduates and post graduate employees – studying in this critical area.
- Provide students with critical scholarship funding
- Provide students with an internship or co-op with a premiere industry partner

The Need

Numerous studies underscore the need for workers capable of using the methodologies of data science.

• McKinsey Global Institute predicts a **nationwide shortage of 140,000 to 190,000 workers** with "deep analytical skills," and **a deficit of 1.5 million managers** capable of using big data analytics for actionable insights in their decision-making.

• McKinsey predicts a **40 percent annual growth in global data** and $**300 billion in potential value-add** of data analytics to the nation's health care industry alone.

• Gartner Inc. indicates that by 2015, **4.4 million information technology jobs** globally will be created to support data science and analytics, generating **1.9 million IT jobs** in the United States.

• The U.S. government anticipates a **shortage of about 50,000 qualified workers** in health IT.

About Case Western Reserve

Case Western Reserve University is one of the country's leading private research institutions. Located in Cleveland, the university offers a unique combination of forward-thinking educational opportunities in an inspiring cultural setting. The renowned faculty engages in teaching and research in a collaborative, hands-on environment. Case Western's nationally recognized programs include arts and sciences, dental medicine, engineering, law, management, medicine, nursing and social work. About 4,200 undergraduate and 5,600 graduate students comprise the student body.

BHEF WINTER 2014 MEETING

DATA SCIENCE OVERVIEW

BHEF BUSINESS-HIGHER EDUCATION FORUM
Creating Solutions. Inspiring Action.

Introduction

BHEF's National Higher Education and Workforce Initiative has served as a powerful platform for launching innovation in regional workforce projects through strategic collaboration between business and higher education. The success of the Initiative since its launch provides a foundation for expanding its focus into an additional emerging field: data science. During BHEF's winter meeting, Case Western Reserve University (CWRU), in collaboration with corporate partners, will announce the launch of a distinctive undergraduate degree program in data science.

The following describes the emergence of data science as an essential tool for decision making and innovation in a wide range of organizations, outlines the demand for data science experts and data analytics-enabled graduates, reports on analysis of existing data science programs at sampled colleges and universities, and highlights work underway on BHEF academic members' campuses to meet the growing workforce needs of business, government, scientific and research organizations, and cultural institutions.

Emerging Workforce Needs in Data Science

The application of data science is pervasive in both the public and private sectors. Companies of all sizes rely on data science and analytics as key transformational components to their core operations. A 2011 report by the McKinsey Global Institute, *Big Data: The Next Frontier for Innovation, Competition, and Productivity*, noted that "big data" is growing at a rate of 40 percent each year and has the potential to add $300 billion of value to the nation's health care industry alone, with broad application in virtually every sector, as well as scientific organizations and cultural institutions. Projections by Gartner, Inc., indicate that in less than 12 months, 4.4 million information technology (IT) jobs to support big data will be created globally. About 1.9 million of those jobs will be within the United States, and big data has the potential to create three times that number of jobs outside of IT. Despite this demand, the U.S. faces a significant shortfall in the number of data scientists and "data-enabled" professionals. According to the McKinsey report, the United States will need an additional 140,000 to 190,000 data science experts with "deep analytical skills," plus 1.5 million managers capable of using data analytics in decision making.

Data science provides new sources of actionable insights that will improve decision making and stoke innovation. While business is able to collect and store vast amounts of data, most struggle to harness these data for decision making. Data becomes business intelligence—and valuable to decision makers—when data science experts have domain-specific expertise to access, unlock, and interpret the data available to them.

Data analytics-enabled individuals—those who understand the processes and tools of data science—with domain-specific expertise can turn data into information, and they are critical to the ability of businesses to implement data-driven decision-making throughout organizations. While there is considerable need nationally for data scientists, there is an even greater workforce need for the analytics-enabled professional who can marry a deep background in a particular field (e.g., engineering, economics, or business) with a strong understanding of the application of analytics and visualization tools.

In addition, the data science "footprint" has expanded from core IT/computer science functions to each business unit and the many functions within these units, including operations, marketing, and communications. Businesses that are able to successfully integrate data analytics-enabled professionals into these functions have gained significant competitive advantages. Yet, this expansion in demand results in a talent gap in data analytics skills in the current and future workforce.

The Challenge

While higher education is responding to the demand from businesses for more data science professionals, this demand quickly outstrips the supply. Today, most higher education data science programs are located within the school of engineering and/or the department of computer science. These programs serve as post-baccalaureate training for individuals who already have a strong grasp of analytical thinking, applied mathematics, and competency in computer programming, largely limiting access to graduates from other STEM fields. Consequently, these graduate programs prepare data science experts who will lack the domain-specific knowledge in health, transportation, economics, business, and public policy that companies need.

To date, very few programs have been developed to offer undergraduates training in data science. Even at higher education institutions with numerous course offerings in data science, undergraduates have limited access to these courses before the final two years of college, and these students will largely be STEM majors. But as a result of high attrition among STEM undergraduates in the first two years, few STEM graduates will actually be exposed to data science. Even fewer non-STEM majors will be exposed to data analytics because data science and analytics courses are not integrated into most undergraduate programs.

To ensure a robust and diverse data science and analytics-enabled workforce, learning opportunities in data science must be integrated into courses across the undergraduate curricula, beginning in the first year.

Strategic Business Engagement Offers a Solution to Undergraduate Pathways in Data Science

Together, business and higher education can address our nation's potential data science workforce deficit. Through deep collaboration with businesses in a variety of industry sectors, higher education can respond to clearly articulated corporate needs, creating curricula and learning opportunities that will build robust undergraduate pathways and produce both data science experts and analytics-enabled graduates across all sectors. Furthermore, collaborations in data science between business and higher education have the potential to create new majors as well as integrate learning opportunities in data science and analytics into the broader higher education curriculum (e.g., the social sciences and humanities) through minors and certificates. Such strategic partnerships could result in a significantly more diverse talent pool entering the data science and analytics-enabled workforce.

BHEF's Initiative provides a platform to develop strategic partnerships and form a network of regional projects led by business and academic members which will create new undergraduate pathways in data science and analytics. BHEF's winter meeting represents a unique opportunity to launch BHEF's inaugural project in data science. To address this need for undergraduates to be prepared with data science skills and knowledge, CWRU will launch a distinctive undergraduate degree program in data science in the fall of 2014. Based on a core curriculum of data-science fundamentals, the CWRU program will prepare graduates to meet business needs through cutting-edge instruction; exciting areas of concentration in health, energy, and manufacturing and production; and an extensive capstone experience. Students completing the data science programs will be well prepared to fill entry-level workforce needs for data scientists in a variety of industries. CWRU will continue its focus on data science in the future through the development of a post-baccalaureate certificate program for business personnel interested in retraining.

Business is partnering with universities to develop opportunities for students to obtain data science skills. The Humana Foundation provided Bellarmine University in Louisville, Kentucky, a multi-year, $1 million grant to support a new program, the Institute for Advanced Analytics. Opening next fall, the institute will offer new informatics and technology degrees at the undergraduate, graduate, and continuing education levels and will develop interdisciplinary data analysis programs in nursing and health sciences, communication, education, business, and environmental studies.

In addition, strategic engagement between business and higher education can support the development of freshman introductory courses to expose students to data science and analytics. Introductory courses could include information that would explain to students how data science may impact their majors and courses of study. Equipped at the very beginning of their undergraduate experience with this knowledge and with business support—to include mentors, internships, and other resources—these students may persist and thus significantly boost the data science pipeline.

Current introductory course redesign efforts across the country can serve as examples for new approaches to expose students from many disciplines to data science and analytics. CWRU's freshman coursework in engineering and materials science, developed as one of BHEF's first regional

projects, is one successful example. Through a series of courses, CWRU first-year students interested in these fields are directly exposed to processes and projects at a diverse array of local medical, aerospace, electric, and advanced-materials technology businesses. Internships and other on-site opportunities supplement project-based teamwork on real-world problems, with students ultimately producing real parts, components, or devices. From the onset of their undergraduate studies, materials science and engineering students are engaged with real-world examples in their chosen field.

For example, as part of the Association of American Universities (AAU) STEM education initiative, the University of Arizona is redesigning introductory STEM courses to improve information and quantitative literacy by incorporating the use of real-life applications in problem-solving. It is also expanding its use of hands-on demonstrations and experiments in introductory courses to further develop conceptual understandings of central theories in biology and engineering. As part of that same AAU initiative, the University of Pennsylvania seeks to improve STEM education for its students by creating blended introductory courses in math, chemistry, physics, and engineering. These are two of several evidence-based approaches that integrate data science into introductory undergraduate courses.

This type of strategic engagement between business and higher education is an excellent strategy to apply to data science, which requires the application of theoretical knowledge to real-world problems for truly relevant learning opportunities for both data science experts and analytics-enabled graduates.

Next Steps: Building the Foundation for a National Data Science Network

Working at the nexus of business and higher education, BHEF is taking a lead role in mapping and executing a robust strategy to help the nation meet its needs for a skilled data science workforce. Intentionally moving beyond engagements that are merely transactional, BHEF plans to engage companies, higher education institutions, government agencies, philanthropies, cultural institutions, and other key stakeholders in coordinated, long-term, action-oriented partnerships. In data science, a discipline that did not exist even a decade ago, BHEF will play a pivotal role in the definition, design, and delivery of powerful new pathways for preparing and educating the data science workforce, including both professionals and enabled employees, particularly at the undergraduate level.

Based on the successful example of the National Cybersecurity Network, launched in 2012 at the Alfred P. Sloan Foundation, BHEF continues to convene top leaders from business, higher education, and state and federal government agencies, with the most recent meeting being held in November 2013 at the IBM-Almaden Research Center in San José, California. The BHEF National Data Science Network will be a forum through which business, higher education, and government share ideas, collaborate to strengthen and diversify the data science workforce, and align undergraduate education in data science with regional workforce requirements. This can be achieved by focusing on the earliest years of the undergraduate experience and increasing the persistence of students to degree completion in this high-demand field.

Data Science for Corporate Decision-Making: Preparing Undergraduates to Unleash the Power of Data

The volume of data acquired or generated by corporations continues to increase at unprecedented rates, and businesses increasingly are working to interpret this data for strategic decision-making. What can data reveal about customer preferences? Production flow? Opportunities for cost savings? The potential impact of data analytics on the way industry does business is akin to the way the internet has shifted paradigms for commerce.

Case Western Reserve University (CWRU) is launching a distinctive undergraduate degree program to fill this workforce need.

Rooted in a core curriculum of data-science fundamentals, the program will prepare graduates to meet industry needs through cutting-edge instruction, exciting areas of concentration, and an extensive capstone experience. Students will receive a strong foundation in topics fundamental to data science (such as mathematics, statistics, and computing algorithms), breadth in allied technical areas (such as biology, physics, or engineering), and a focus on a specific domain area: health, energy, and manufacturing and production. As the program matures, other concentrations such as finance may be added.

All of the data science programs will include a strong experiential learning component. Students are also expected to complete a minimum of two internships in data science (one in the summer between sophomore and junior year, the second in the summer between junior and senior year) or one seven-month, full-time, co-op assignment. In addition, all students must complete a two-semester senior capstone project in data science.

Students completing the data science programs will be well prepared to fill entry-level workforce needs for data scientists in a variety of industries. CWRU also will launch a post-baccalaureate certificate program for industry personnel interested in retraining.

These new programs grew from a data science industry roundtable hosted by CWRU last fall. Business leaders discussed their workforce needs in data science skills, their eagerness to hire both recent and mid-career graduates, and the importance for these graduates to come with real-world experience. CWRU's new data science program will reflect their feedback and thrive with their engagement and support.

Field	Data Science Application	Resulting Outcome
Agronomy	Sensors were added to John Deere equipment to help farmers manage their fleets' usage. The information is combined with historical and real-time data regarding weather prediction, soil conditions, and crop features.	Farmers maximized return on crops by implementing recommended crop selection, location, and timing, as well as plowing technique, methodology, and location.[1]
Archeology	A team of doctors and archeologists analyzed data from whole-body CT scans of 137 mummies spanning 4,000 years.	The team demonstrated that humans have been affected by atherosclerosis since preindustrial times, proving it is not a modern-day disease.[2]
Astronomy	Four years of data from NASA's orbiting Kepler telescope were analyzed to compute how many planets lie in their solar systems' "Goldilocks zone," where surface temperatures support liquid water.	Astronomers calculated 11 billion potentially habitable planets in our galaxy. One in five sun-like stars harbors a roughly Earth-size planet in the habitable zone, with the nearest possibly close enough for communication.[3]
Biology	ENCyclopedia Of DNA Elements (ENCODE) is an open research consortium aiming to rigorously analyze the human genome sequence and identify all functional elements, complementing the completed Human Genome Project.	ENCODE provides insight into how genetic variations affect human traits and diseases by assigning a function to approximately 80% of the human genome.[4]
Climatology	Weather reports, tidal phases, satellite images, and deforestation maps were analyzed to improve siting of wind turbines by Vestas, a global energy company.	These efforts increased turbine productivity by reducing response time to implement wind forecasting information by 97%.[5]

1 http://www.bigdata-startups.com/BigData-startup/john-deere-revolutionizing-farming-big-data/.

2 Thompson R.C., Allam A.H., Lombardi G.P., et al. (2013). Atherosclerosis across 4000 years of human history: the Horus study of four ancient populations. *The Lancet* 381: 1211-1222.

3 Petigura E., Howard A., and Marcy G. (2013). Prevalence of Earth-size planets orbiting Sun-like stars. *Proceedings of the National Academy of Sciences of the United States of America* 110: 19273-19278.

4 Maher B. (2012). ENCODE: The human encyclopedia. *Nature* 489: 46-48.

5 "Vestas Wind Systems Turns to IBM Big Data Analytics for Smarter Wind Energy." IBM News Release (October 24, 2011).

Field	Data Science Application	Resulting Outcome
Criminology	Social network analysis was used on a data set of 60 active gangs and 600 factions by the Chicago Police Department to determine social and geographic factors that connect and divide gang members.	Police reduced, by 20%, city-wide homicides by using analysis results to predict future potential victims of gang-related violence and monitoring those areas.[6]
Economics	Housing-related Google search queries were analyzed to determine alternate methods of real-estate forecasting.	Trends of increasing or decreasing volumes of Google queries were more accurate predictors of future house sales than expert forecasts by real estate economists.[7]
Education Policy	Test scores, attendance records, tax returns, and demographic information on 2.5 million New York City schoolchildren were analyzed and correlated with their adult earnings to study the long-term effects of teaching quality, as determined by analysis of administrative performance records.	The study revealed that a student's lifetime earnings increased by one-quarter of a million dollars in present value terms by replacing a poorly rated teacher with an average-rated teacher.[8]
Engineering	During missile production, data such as the number of turns on a screw in assembly operations are continuously collected and evaluated by Raytheon. To determine equipment at risk for failure, predictive analysis is applied to temperature, acoustic, and visual data collected from Union Pacific rail cars, including noises from wheel vibrations and ultrasound wheel visuals.	Error messages report anomalies in missile assembly, halting production and reducing defects in real-time, ultimately reducing production costs.[9] Spending reduced by millions as a result of a 75% reduction in bearing-related derailments.[10]

[6] Tony Dokoupil. "'Small world of murder': As homicides drop, Chicago police focus on social networks of gangs." *NBC News* (December 17, 2013). http://usnews.nbcnews.com/_news/2013/12/17/21864122-small-world-of-murder-as-homicides-drop-chicago-police-focus-on-social-networks-of-gangs?lite.

[7] Wu L. and Brynjolfsson E. (2009). The Future of Prediction: How Google Searches Foreshadow Housing Prices and Quantities. *ICIS 2009 Proceedings. Paper* 147.

[8] Chetty R., Friedman J., and Rockoff J. (2011). The Long-Term Impacts of Teachers: Teacher Value-Added and Student Outcomes in Adulthood. NBER Working Paper No. 17699.

[9] Ahmed Noor. "Putting Big Data to Work." https://www.asme.org/wwwasmeorg/media/ResourceFiles/Network/Media/Mechanical%20Engineering%20Magazine/1013BigData.pdf.

[10] Chris Murphy. "Union Pacific Delivers Internet of Things to Reality Check." Information Week. August 3, 2012. http://www.informationweek.com/it-leadership/union-pacific-delivers-internet-of-things-reality-check/d/d-id/1105644?.

Field	Data Science Application	Resulting Outcome
Finance	Pioneer West Virginia Credit Union applied data analytics to its loan and deposits portfolios, with a focus on loan delinquencies.	Daily delinquency reports replaced monthly delinquency reports, allowing for early intervention, which decreased the loan delinquency rate by 110 basis points within 5 months.[11]
Literary History	Key words, phrases, and linguistic patterns from 3,592 literary works published from 1780 to 1900 were computationally analyzed.	Analysis revealed Jane Austen and Sir Walter Scott had the greatest influence on authors' writing style and themes.[12]
Literature	A Google database of 15 million scanned books spanning several centuries was analyzed for fluctuations in word choice.	Year-to-year fluctuations in the use of particular words allowed the researchers to examine the evolution of the novel in the United States and changes in American society and values.[13]
Political Science	Barack Obama's 2012 re-election campaign staff built predictive models of voter behavior using demographic information such as age, sex, race, neighborhood, voting records, and consumer data.	Using their analyses, the Obama campaign raised more than $1 billion by increasing the efficacy of traditional fundraising strategies, such as phone calls, direct mailings, and social media, and increased Obama voter turnout by more efficiently placing TV ads to target swing-state voters.[14]
Psychology	Demographic and linguistic data from Facebook were combined with responses from personality questionnaires by University of Pennsylvania researchers.	Novel insights into the relationship between personal traits and language were recognized, such as a correlation between emotional stability and sports references.[15]

[11] "Pioneer accelerates insight and enhances decision-making." IBM Case Study (June 9, 2011). http://www-01.ibm.com/software/success/cssdb.nsf/CS/STRD-8HKM2P?OpenDocument&Site=default&cty=en_us.

[12] Matthew L. Jocklers. *Macroanalysis: Digital Methods and Literary History*. University of Illinois Press, 2013.

[13] Marc Egnal. Evolution of the Novel in the United States: The Statistical Evidence. *Social Science History* (2013) 37:2. http://muse.jhu.edu/login?auth=0&type=summary&url=/journals/social_science_history/v037/37.2.egnal.html.

[14] Michael Schere. Inside the Secret World of the Data Crunchers Who Helped Obama Win. *Time* (November 7, 2012).

[15] Penn Researchers Use Facebook Data to Predict Users' Age, Gender and Personality Traits. http://www.upenn.edu/pennnews/news/penn-researchers-use-facebook-data-predict-users-age-gender-and-personality-traits.

BHEF BUSINESS-HIGHER EDUCATION FORUM®

Creating Solutions. Inspiring Action.

WINTER MEMBER MEETING

February 6-7, 2014

Hotel Monaco Gallup
Headquarters
Washington, D.C.

Key Word Glossary: Data Science & Analytics

Data Sources ⇔ Informatics ⇔ Analytics ⇔ Actionable Intelligence

Analytics – the discovery of insights in data

Data Science –builds on techniques and theories from many fields, including mathematics, statistics, and computer science related topics such as data engineering, metadata and terminology systems, pattern recognition and machine learning, data mining, advanced algorithms, user interfaces and visualization, privacy and security, uncertainty modeling, data storage and management, and high performance computing with the goal of extracting meaning from data and creating data products.

Data source – a collection or a stream of data such as a Twitter feed or weblog

Exploratory data analysis (EDA) – finding patterns within data before applying a standard or new analytics tool. It is a means of exploring the data and to find the main characteristics.

Cloud computing – a framework such as Hadoop that is built to enable the process and storage of big data across a distributed file system

Internet of Things – ordinary devices that are connected to the internet at any time any via sensors

Open Access Science – a US OSTP policy to make the published results of federally funded research freely available

Project Open Data – A US OSTP initiative. Project Open Data is an online, public repository

Statistics – is the study of the collection, organization, analysis, interpretation and presentation of data.

Full Glossary: Data Science & Analytics

Data Sources ⇔ Informatics ⇔ Analytics ⇔ Actionable Intelligence

A

Aggregation – a process of searching, gathering and presenting data
Algorithms – a mathematical formula that can perform certain analyses on data
Analytics – the discovery of insights in data
Anomaly detection – the search for data items in a dataset that do not match a projected pattern or expected behavior. Anomalies are also called outliers, exceptions, surprises or contaminants and they often provide critical and actionable information.
Anonymization – making data anonymous; removing all data points that could lead to identify a person

B

Big Data Scientist – someone who is able to develop the algorithms to make sense out of big data
Business Intelligence – the theories, methodologies and processes to make data understandable

C

Classification analysis – a systematic process for obtaining important and relevant information about data, also meta data called; data about data.
Cloud computing – a distributed computing system over a network used for storing data off-premises
Clustering analysis – the process of identifying objects that are similar to each other and cluster them in order to understand the differences as well as the similarities within the data.
Comparative analysis – it ensures a step-by-step procedure of comparisons and calculations to detect patterns within very large data sets.
Complex structured data – data that are composed of two or more complex, complicated, and interrelated parts that cannot be easily interpreted by structured query languages and tools.
Correlation analysis – the analysis of data to determine a relationship between variables and whether that relationship is negative (- 1.00) or positive (+1.00).

D

Data aggregation tools - the process of transforming scattered data from numerous sources into a single new one.
Data analyst – someone analyzing, modeling, cleaning or processing data
Data Analysis – Analysis of data is a process of inspecting, cleaning, transforming, and modeling data with the goal of discovering useful information, suggesting conclusions, and supporting decision making.
Database – a digital collection of data stored via a certain technique
Database-as-a-Service – a database hosted in the cloud on a pay per use basis, for example Amazon Web Services
Database Management System – collecting, storing and providing access of data
Data cleaning – the process of reviewing and revising data in order to delete duplicates, correct errors and provide consistency
Data ethical guidelines – guidelines that help organizations being transparent with their data, ensuring simplicity, security and privacy
Data source – a stream of data such as a Twitter feed or RSS
Data mining – the process of finding certain patterns or information from data sets
Data modeling – the analysis of data objects using data modeling techniques to create insights from the data
Data set– a **collection of data**
De-identification – same as anonymization; ensuring a person cannot be identified through the data

E

Exploratory data analysis (EDA) – finding patterns within data without standard procedures or methods. It is a means of discovering the data and to find the data sets main characteristics.
Extract, Transform and Load (ETL) – a process in a database and data warehousing meaning extracting the data from various sources, transforming it to fit operational needs and loading it into the database

G

Gamification – using game elements in a non-game context; very useful to create data therefore coined as the friendly scout of big data

Graph Analysis – viewing relationships among the nodes in terms of the network or graph theory, meaning analyzing connections between nodes in a network and the strength of the ties.

Graph Databases – they use graph structures (a finite set of ordered pairs or certain entities), with edges, properties and nodes for data storage. It provides index-free adjacency, meaning that every element is directly linked to its neighbor element.

H

Hadoop – an open-source framework that is built to enable the process and storage of big data across a distributed file system

HBase –an open source, non-relational, distributed database running in conjunction with Hadoop

HDFS – Hadoop Distributed File System; a distributed file system designed to run on commodity hardware

High-Performance-Computing (HPC) – using supercomputers to solve highly complex and advanced computing problems

I

Internet of Things – ordinary devices that are connected to the internet at any time any via sensors

J

Juridical data compliance – relevant when you use cloud solutions and where the data is stored in a different country or continent. Be aware that data stored in a different country has to oblige to the law in that country.

K

KeyValue Databases – they store data with a primary key, a uniquely identifiable record, which makes easy and fast to look up. The data stored in a KeyValue is normally some kind of primitive of the programming language.

L

Linked Data – Linked Data describes a method of publishing structured data so that it can be interlinked and become more useful. It builds upon standard Web technologies such as HTTP, RDF and URIs

Load balancing – distributing workload across multiple computers or servers in order to achieve optimal results and utilization of the system

Location data – GPS data describing a geographical location

M

Machine data – data created by machines via sensors or algorithms

Machine learning – part of artificial intelligence where machines learn from what they are doing and become better over time

MapReduce – a software framework for processing vast amounts of data

Massive Data Analysis – A 2013 National Research Council report on how data mining of massive data sets is transformative. Massive datasets are potential sources of discovery and knowledge, requiring sophisticated analysis techniques aiming to find relational and semantic interpretations of the phenomena underlying the data. http://www.nap.edu/catalog.php?record_id=18374

Massively Parallel Processing (MPP) – using many different processors (or computers) to perform certain computational tasks at the same time

Metadata – data about data; gives information about what the data is about.

MongoDB – an open-source NoSQL database

Multi-Dimensional Databases – a database optimized for data online analytical processing (OLAP) applications and for data warehousing.

MultiValue Databases – they are a type of NoSQL and multidimensional databases that understand 3 dimensional data directly. They are primarily giant strings that are perfect for manipulating HTML and XML strings directly

N

Natural Language Processing – a field of computer science involved with interactions between computers and human languages

Network analysis – viewing relationships among the nodes in terms of the network or graph theory, meaning analyzing connections between nodes in a network and the strength of the ties.

NewSQL – an elegant, well-defined database system that is easier to learn and better than SQL. It is even newer than NoSQL

NoSQL – sometimes referred to as 'Not only SQL' as it is a database that doesn't adhere to traditional relational database structures. It is more consistent and can achieve higher availability and horizontal scaling.

O

Object Databases – they store data in the form of objects, as used by object-oriented programming. They are different from relational or graph databases and most of them offer a query language that allows object to be found with a declarative programming approach.

Object-based Image Analysis – analyzing digital images can be performed with data from individual pixels, whereas object-based image analysis uses data from a selection of related pixels, called objects or image objects.

Open Access Science – a US OSTP policy to make the published results of federally funded research freely available to the public within one year of publication and requiring researchers to better account for and manage the digital data resulting from federally funded scientific research. http://www.whitehouse.gov/blog/2013/02/22/expanding-public-access-results-federally-funded-research

Open Data Charter – a policy of the G8 nations, singed in 2013. The Open Data Charter sets out 5 strategic principles including an expectation that all government data will be published openly by default, alongside principles to increase the quality, quantity and re-use of the data that is released. G8 members have also identified 14 high-value areas – from education to transport, and from health to crime and justice – from which they will release data. https://www.gov.uk/government/publications/open-data-charter

Open Data Policy – US OSTP Policy on Open Data http://www.whitehouse.gov/sites/default/files/omb/memoranda/2013/m-13-13.pdf

Openness – the rise of the Internet and the digitization of information are increasing the "openness" of information, processes, and institutions. http://www.ced.org/reports/single/harnessing-openness-to-improve-research-teaching-and-learning-in-higher-edu

Open Source – a development model to promote universal access via free license to a product's design or blueprint, and universal redistribution of that design or blueprint, including subsequent improvements to it by anyone.

Open Source Software – computer software with its source code made available and licensed with a license in which the copyright holder provides the rights to study, change and distribute the software to anyone and for any purpose

Optimization analysis - the process of optimization during the design cycle of products done by algorithms. It allows companies to virtually design many different variations of a product and to test that product against pre-set variables.

Ontology – ontology represents knowledge as a set of concepts within a domain and the relationships between those concepts

Outlier detection – an outlier is an object that deviates significantly from the general average within a dataset or a combination of data. It is numerically distant from the rest of the data and therefore, the outlier indicates that something is going on and generally therefore requires additional analysis.

P

Pattern Recognition – identifying patterns in data via algorithms to make predictions of new data coming from the same source.

Predictive analysis/analytics – the most valuable analysis/analytics within big data as they help predict what someone is likely to buy, visit, do or how someone will behave in the (near) future. It uses a variety of different data sets such as historical, transactional, social or customer profile data to identify risks and opportunities.

Privacy – to seclude certain data / information about oneself that is deemed personal

Project Open Data – A US OSTP initiative. Project Open Data is an online, public repository intended to foster collaboration and promote the continual improvement of the Open Data Policy. http://www.whitehouse.gov/blog/2013/05/16/introducing-project-open-data Project Open Data is hosted on GitHub http://project-open-data.github.io/

Public data – public information or data sets that were created with public funding

Q

Query – asking for information to answer a certain question

R

Real-time data – data that is created, processed, stored, analyzed and visualized within milliseconds

Recommendation engine – an algorithm that suggests certain products based on previous buying behavior or buying behavior of others

Re-identification – combining several data sets to find a certain person within anonym zed data

Regression analysis – to define the dependency between variables. It assumes a one-way causal effect from one variable to the response of another variable.

RFID – Radio Frequency Identification; a type of sensor using wireless non-contact radio-frequency electromagnetic fields to transfer data

Reproducible Science – An effort of the US NSF to advance the idea that the ultimate product of academic research is the paper along with the full computational environment used to produce the results in the paper such as the code, data, etc. that can be used to reproduce the results and create new work based on the research. http://www.reproduciblescience.org/index.php/Main_Page

Resource Description Framework (RDF) – a metadata data model which is part of the semantic web effort. In the form of subject-predicate-object expressions. These expressions are known as triples in RDF terminology

S

Semantic Web – a project of W3C, which "provides a common framework that allows data to be shared and reused across application, enterprise, and community boundaries

Semi-structured data - a form a structured data that does not have a formal structure like structured data. It does however have tags or other markers to enforce hierarchy of records.

Signal analysis – it refers to the analysis of measurement of time varying or spatially varying physical quantities to analyze the performance of a product. Especially used with sensor data.

Simulation analysis – a simulation is the imitation of the operation of a real-world process or system. A simulation analysis helps to ensure optimal product performance taking into account many different variables.

Smart grid – refers to using sensors within an energy grid to monitor what is going on in real-time helping to increase efficiency

Software-as-a-Service – a software tool that is used of the web via a browser

Spatial analysis – refers to analyzing spatial data such geographic data or topological data to identify and understand patterns and regularities within data distributed in geographic space.

Statistics – is the study of the collection, organization, analysis, interpretation and presentation of data. It deals with all aspects of data, including the planning of data collection in terms of the design of surveys and experiments.

SQL – a programming language for retrieving data from a relational database

Structured data – data that is identifiable as it is organized in structure like rows and columns. The data resides in fixed fields within a record or file or the data is tagged correctly and can be accurately identified.

T

Time series analysis - analyzing well-defined data obtained through repeated measurements of time. The data has to be well defined and measured at successive points in time spaced at identical time intervals.

Topological Data Analysis – focusing on the shape of complex data and identifying clusters and any statistical significance that is present within that data.

Transactional data – dynamic data that changes over time

Transparency – consumers want to know what happens with their data and organizations have to be transparent about that

U

Un-structured data - unstructured data is regarded as data that is in general text heavy, but may also contain dates, numbers and facts.

V

Value – all that available data will create a lot of value for organizations, societies and consumers. Big data means big business and every industry will reap the benefits from big data.

Variability – it means that the meaning of the data can change (rapidly). In (almost) the same tweets for example a word can have a totally different meaning

Variety – data today comes in many different formats: structured data, semi-structured data, unstructured data and even complex structured data

Velocity – the speed at which the data is created, stored, analyzed and visualized

Veracity – organizations need to ensure that the data is correct as well as the analyses performed on the data are correct. Veracity refers to the correctness of the data

Visualization – with the right visualizations, raw data can be put to use. Visualizations of course do not mean ordinary graphs or pie-charts. They mean complex graphs that can include many variables of data while still remaining understandable and readable

Volume – the amount of data, ranging from megabytes to brontobytes

W

Weather data – an important open public data source that can provide organizations with a lot of insights if combined with other sources

X

XML – Extensible Markup Language (XML) is a markup language that defines a set of rules for encoding documents in a format that is both human-readable and machine-readable.

Glossary information from http://www.bigdata-startups.com/abc-big-data-glossary-terminology/, Wikipedia, US OSTP, G8 - gov.uk

EECS/SYBB 459: Bioinformatics for Systems Biology Spring 2013

Course Syllabus[a]

Instructor: Mehmet Koyutürk

[a]Everything in this syllabus is subject to change based on students' suggestions and instructor's improvements.

Course Objectives

In recent years, there has been an explosion in generation of biological data that provide genome-scale information on cellular processes. These data include genomic sequences, gene expression, protein expression, protein-protein interactions, protein-DNA interactions, metabolite concentrations, and metabolic fluxes, among others. Such data are often analyzed within the framework of biological network models, which provide a large-scale view of the functional relationships among multiple molecules in the cell.

This course provides a detailed review of computational approaches that are developed to analyze these novel sources of biological data. Topics include prediction of protein-protein interactions, genetic regulatory network inference, analysis of network topology, network motifs, network alignment, network-based functional annotation, gene set enrichment analysis, metabolic flux analysis and network-based analysis of genomic, functional genomic, and proteomic data with applications to the systems biology of complex diseases. It is expected that, upon completion of this course, the students will achieve the following objectives:

- Become familiar with existing tools and resources for computational analysis of "omic" datasets.
- Develop an awareness of the computational problems that arise in the modeling and analysis of cellular systems.
- Understand fundamental abstractions and computational approaches used to formulate and address these problems.
- Be able to use, manipulate, and extend existing computational infrastructure for analyzing systems biology data.

Class Meeting

MW 1:45-3:00 PM, Sears 354.

Instructor

- Mehmet Koyutürk
  Office: Olin 512
  Phone: 368-2963
  e-mail: `mxk331@case.edu`
  Office hours: MW 11:00 AM - 12: 00PM and by appointment.

Course Work & Grading

Assignments: (30%) There are three practical assignments. In each assignment, one or more datasets will be provided, and the students will be asked to analyze the data to answer specific questions, possibly using publicly available computational resources and public data.

Paper Discussion: (Leading 20%, Contribution 10%) We will discuss several research papers on topics that are covered in the class. In each discussion session, we will discuss one paper. Each Monday, one student will present a recent research paper and lead the discussion on this paper. Each student in the class will review one of the papers and submit their review of the research papers to the corresponding presenter by 5 PM on Friday. During the class, the presenters will summarize the paper (key ideas, main contributions, experimental results, conclusions) and provide a discussion of the reviews submitted by all students. Based on this content, the presenters will arbitrate a discussion on the intellectual merit and potential impact of these papers, their limitations, and ideas for building on their results.

Project: (15% Proposal, 25% Final Presentation and Report) The students will develop and conduct research projects that make use of omic data. The projects will be conducted in teams of two and **the teams are required to be interdisciplinary.** The projects are expected to be (i) innovative, (ii) involve implementation or in-depth use of a computational method, and (iii) provide solid results. The development of projects comprises the following phases:

1. *Topic selection.* Students will form teams and select their topic by Wednesday, February 12.

2. *Project proposal.* The teams will review the literature and develop ideas for their project. Based on these ideas, they will present a proposal for their project on Monday/Wednesday, March 17/19. Other students and the instructor will provide feedback on these ideas, critically evaluating the plan to execute these ideas. The presentation will not be graded.

   Based on this feedback, the students will submit a *5-page research proposal* by Wednesday, March 26. The proposal will outline the motivation and the proposed idea and clearly argue for the significance and the intellectual merit of the proposed research. It should also clearly explain the research plan. The proposal will be expected to address the feedback provided during the presentations and will be graded.

3. *Project presentation and report.* On the final week of the semester, each team will present their project. Each team will also submit an *8-page project report* by Wednesday, April 30. The presentations and reports will be graded.

Calendar

The class meetings are organized as follows: Every week (a Wednesday followed by a Monday) a different topic will be discussed. On Wednesday, the instructor will provide background on the topic. Prior to Wednesday's class meeting, review articles or book chapters on the topic will be posted on Blackboard. On the following Monday, a research paper will be discussed, and the discussion will be facilitated by a student. The topics scheduled for discussion are the following:

Date	Topic
Jan 13	Introduction to "omics"
Jan 15–22	Data on biological function and relationship among biomolecules: Signaling, pathways, protein-protein interactions, functional annotation, data and knowledge bases
Jan 27	*Paper Discussion #1:* Dutkowski et al. A gene ontology inferred from molecular networks, Nat Biotech, 2013.
Jan 29	Network analysis, graph theoretical concepts, topological properties of biological networks
Feb 3	*Paper Discussion #2:* Heo et al., Topology of protein interaction network shapes protein abundances and strengths of their functional and nonspecific interactions, PNAS, 2011.
Feb 5	Network motifs, module identification, network alignment, network evolution
Feb 10	*Paper Discussion #3:* Mazurie et al. An evolutionary and functional assessment of regulatory network motifs, Genome Biol, 2005.
Feb 12	Centrality, network based functional annotation, genome-wide linkage and association, disease gene prioritization
Feb 17	*Paper Discussion #4:* Lee et al. Prioritizing candidate disease genes by network-based boosting of genome-wide association data, Genome Res, 2012.
Feb 19	Gene expression: clustering, classification, regulatory network inference
Feb 24	*Paper Discussion #5:* Sirota et al. Discovery and preclinical validation of drug indications using compendia of public gene expression data, Science Transl Med, 2011.
Feb 26 – Mar 3	Cancer: Driver mutations, methylation, mRNA-level dysregulation, proteomics, The Cancer Genome Atlas (TCGA), prediction of patient outcome
Mar 5	*Paper Discussion #6:* Ciriello et al. Mutual exclusivity analysis identifies oncogenic network modules., Genome Res, 2012.
Mar 17–19	Proposal Presentations
Mar 24	Information theory and its applications in biology
Mar 26	Active pathways, dysregulated pathways, Gene-set enrichment analysis,
Mar 31	*Paper Discussion #7:* Weng et al. SNP-based pathway enrichment analysis for genome-wide association studies, BMC Bioinformatics, 2011.
Apr 2	Dysregulated subnetworks, composite gene features for classification
Apr 7	*Paper Discussion #8:* Patel et al. Network signatures of survival in glioblastoma multiforme, PLoS Comp Biol, 2013.
Apr 9	Metabolomics, metabolic networks, flux-balance analysis, metabolic pathway databases
Apr 14	*Paper Discussion #9:* Zelezniak et al. Metabolic network topology reveals transcriptional regulatory signatures of type 2 diabetes., PLoS Comp Biol, 2010.
Apr 16	*Paper Discussion #10:* Hofree et al. Network-based stratification of tumor mutations, Nat Methods, 2013.
Apr 21–23–29	Project Presentations

Plagiarism Policy

All of the materials submitted by the students must be the students' own work. Zero-tolerance policy on plagiarism will be enforced. Following the departmental plagiarism policy, cheating on homeworks or tests will result in an F grade for the whole course and appropriate disciplinary action, independently of the extent of plagiarism. In case of doubt, the students are responsible for checking with the instructor on what is allowed and what is not.

ASTR 306/406 – Astronomical Techniques

This course will focus on research techniques in astronomy, including the acquisition, reduction, and analysis of data from ground- and space-based telescopes, as well as data-mining from online multi-wavelength datasets. We will also emphasize the variety of writing that astronomers do, such as observing proposals, journal articles, technical reviews, and funding proposals. ASTR 306 is an approved SAGES departmental seminar.

Instructor: Chris Mihos (Sears 557, mihos@case.edu)

Textbooks: While there are no required textbooks, readings will come from a variety of sources available online and in the astronomy library, including:

- **BGO**: *Observational Astronomy*, by Birney, Gonzales, and Oesper
- **Chromey**: *To Measure the Sky*, by Chromey
- **Howell**: *Handbook of CCD Astronomy*, by Howell

Computational Requirements: High level computational data analysis will be required. The specific language or software you use for this analysis is your own choice — depending on the scope of the problem, any of Python, C/C++, Matlab, IRAF, or Supermongo may be appropriate. If you do not have an account on the departmental Linux workstations, please see Charley Knox to get set up ASAP.

Individual Assignments: Every few weeks there will be a homework set geared towards the development of technical skills. Typical assignments could include conducting a photometric analysis of astronomical image data, or downloading and analyzing appropriate astronomical datasets off of the web. In addition, there will be several individual writing assignments throughout the semester, which might include

- an observing proposal
- a grant proposal for research using an on-line database
- referee reports of the group projects

Group Projects: Several larger research projects will be done in small groups. For each project, the group must submit a final project report written in the form of an *Astrophysical Journal* paper. The first of these reports will be peer-reviewed by the other groups in the class, after which revisions will need to be made.

ASTR 406 Requirements: Graduate students enrolled in ASTR 406 will have additional problems on the individual assignments that explore topics in more detail. ASTR 406 students will also complete *individually* the group projects assigned for ASTR 306. There will also be an observing trip for ASTR 406 students to Kitt Peak, Arizona to obtain data for an additional data analysis project.

Grading Structure:

	ASTR 306	ASTR 406
Individual Assignments	70%	55%
Projects	30%	30%
Observing Project	--	15%

Course Topics:

I. Astronomical Data
   a. Coordinate systems
   b. Magnitudes and filter systems
   c. Extended objects: shape and structure parameters
   d. Spectroscopic data: spectra and spectral indices

II. Astronomical Instrumentation
   a. Telescopes
   b. Detectors and imaging
   c. Spectrographs

III. Data Reduction and Analysis
   a. Photometry
   b. Spectroscopy
   c. Statistical methods

IV. Multiwavelength Instrumentation and Datasets
   a. Infrared
   b. Radio
   c. Ultraviolet
   d. X-ray

V. Data-Mining in Astronomy
   a. Data products: raw data versus value-added catalogs
   b. Simulation datasets
   c. Online datasets: SDSS, MAST, NED, etc
   d. Data slicing and cross-matching
   e. Data visualization

VI. Case Studies in Astronomical Data Mining
   a. Active galaxies in clusters
   b. Star streams in galaxies
   c. Hunting for exoplanets
   d. Galaxies in the early universe

Case Western Reserve University
The Weatherhead School of Management
MKMR 310, Fall 2013: T 11.30am-2.15pm

Marketing Analytics

Instructor: Jagdip Singh Office Hours: After class, Tuesday, or by appt.
Course Assistant: Hector Martinez (ham48)
Co-Instructors (listed alphabetically): Chris Linderwell & Elisabeth Smith
(All from Rosetta; email: First name.Last name@rosetta.com)
Contact: jagdip.singh@case.edu, PBL 221

Course & Learning Objectives

Develop skills for analyzing marketing data for making intelligent decisions about marketing investments that create value and build competitive advantage. In short, this course will build capabilities for marketing ai—**a**nalytics *for* **i**nsights.

You might be asking, "If I am going to be a marketing professional, do I need hard analytical skills?" The answer: "Absolutely, yes." The divide between soft and hard areas of marketing is a myth. In a digital age, marketing professionals have to be good at both to be taken seriously.

"Do you have to be a math or statistics wiz to be good at marketing analytics?" Answer: "No." Analytics do involve knowledge of basic statistics, math and commercial software (EXCEL & SPSS). These basic skills are a necessary tool-kit for marketing professionals. Being a wiz can help, but is not necessary. At its core, analytics rely on your powers of understanding marketing problems, matching problems to appropriate analytical procedures, and extracting insights from the obtained.

This course should get you in the habit of asking, "What is the underlying problem responsible for visible symptoms in this situation?," "What data do I need to uncover novel solutions to the problem?" and "What are the appropriate analyses for these data to draw insights?" The course will build skills for answering these questions.

In building these skills, the course will use different materials and deliverables with three common learning objectives:

1. Analytics Application (**AA**), applying covered analytical skills to analyze managerial problems with marketing data.
2. Evaluation and Plotting (**EP**), evaluating the results from analytics and plotting those that are useful for problem solving.
3. Integration and Interpretation (**II**), integrating the useful results obtained and developing an interpretation to guide managerial action.

ROSETTA PARTNERSHIP

This course will be delivered in partnership with Rosetta (http://www.rosetta.com/), an interactive consulting company committed to transforming marketing for the connected world. The purpose of this partnership is to:

- Enrich student learning by connecting analytical theory with real-world problem solving.
- Build skills for analytics driven insights that are relevant for tomorrow's marketplace.

To achieve these purposes, the course content and assignments are designed to (a) enable student learning by building ai skills using real-world problems and data, and (b) emphasize analytical tools and software that build students' ai competence giving them a competitive advantage for a range of marketing jobs.

COURSE FORMAT:

Before Each Class

- Read/Review assigned materials. Be prepared to ask questions and get clarifications.
- Review feedback on your assignment. See where you did well, and where you can improve. Bring your questions to class.
- If you have been asked by the instructor to make a class presentation, take the time to develop and sharpen your presentation skills. Present professionally.
- Be prepared for a class quiz.

Beginning of & During Each Class

- Be on time for class. Miss class only in extreme circumstances. Inform instructor in advance. Never miss submitting or doing your assigned work as required per class schedule.
- Be ready to show you are prepared. Ask questions. Pay attention.
- Assignments due that week will be discussed. It is your responsibility to note deadlines and obtain clarifications for completing the assignment.
- Be engaged. Think how you can use the concepts and tools for real life problems.

After Each Class

- Review material covered in class. Note questions, concepts or steps you don't get. Clarify in the next class.
- Re-Review analytics tutorial slides. Attempt at reproducing the results. Start working on the assigned homework and/or extra practice assignment.
- Plan for timely submission of assignment.

Text, Software, Cases and Readings

Text: **SPSS** *(Required)***:** IBM SPSS for Introductory Statistics, Fifth Edition, George Morgan and others, Routledge, ISBN: 978-1-84872-982-7.
EXCEL: No textbook is assigned. Instead, links to readings and videos are provided for each session. All are free.

Required Software: SPSS 20 & Excel. You are ***required*** to download & install from Case Software Center the latest version of SPSS on your own personal computer. Also, you are expected to have the latest version of EXCEL installed on your personal computer. Both are **essential** for the course.

IBM© SPSS Analytics Certification (ACE)

This course prepares you toward sitting in a SPSS certification exam sponsored by IBM. This certification is not a course requirement, but is recommended for students seeking jobs with some analytical component. By itself, this course is not sufficient preparation for this certification exam. Students are advised to take one or more SPSS-based courses that build their skills before sitting for the certification exam. Discuss options with the instructor. See http://www-03.ibm.com/certify/certs/47100101.shtml for more details. The IBM student portal link is https://www.ibm.com/developerworks/community/groups/service/html/communityview?communityUuid=ab443019-1134-4fbb-bc75-eccc1186b261

Pre-requisites and Prepping for the Course

Understanding of basic statistics (as usually covered in the core statistics course) is a pre-requisite. A good work ethic, and a positive learning attitude are keys to strong course performance.

Some prepping suggestions:

1. Review material you covered in the undergraduate statistics course. Online materials are available, usually free, for this purpose (e.g., http://davidmlane.com/hyperstat/; http://onlinestatbook.com/2/index.html).
2. Familiarize yourself with the SPSS statistical software. SPSS has a useful tutorial (under the "Help" tab in the main data screen). Many online resources are available (http://www.hmdc.harvard.edu/projects/SPSS_Tutorial/spsstut.shtml, http://www.ats.ucla.edu/stat/spss/modules/default.htm), and you can download a free "guidebook" by visiting http://www.spsstools.net/ (use the link under "purposes").

Course Structure and Expectations

Attendance and Conduct Policy*:* You are responsible for attending all sessions. In case you have to miss a class for unavoidable reasons, inform the instructor in advance. Absence is excused only under extreme circumstances. No more than 2 excused absences will be permitted. Even for excused absences, the student will be responsible for submitting timely assignments and making up for absences. There will be no make-ups for quizzes. Each student is responsible for proper conduct. Student behavior deemed to be discreditable as determined by the instructor may result in a reduction of grade for the assignment/s involved or for the course as a whole or both.

In short: *Do your work with diligence.*

Academic Responsibility. All students in this course are expected to adhere to university standards of academic integrity. Cheating, plagiarism, and other forms of academic dishonesty will not be tolerated. This includes, but is not limited to: consulting with another person during an exam, turning in written work that was prepared by someone other than you, and making minor modifications to the work of someone else and turning it in as your own. *Unless properly referenced or quoted, submitted work is assumed to be original contribution of the student.* Ignorance will not be permitted as an excuse. If you are not sure whether something you plan to submit would be considered either cheating or plagiarism, it is your responsibility to ask for clarification. When your name appears on an assignment, you are responsible for the integrity of the work, even if you did not personally write the offending material. Information on citations and plagiarism can be found on the following web sites: http://library.case.edu/ksl/researchtools/citation/index.html and http://www.indiana.edu/~wts/pamphlets/plagiarism.shtml. The Weatherhead Academic Integrity policy can be found at http://bulletin.case.edu/weatherheadschoolofmanagement/policies/#academicintegritypolicytext.

Evaluation, Grading and Conduct

Homework Exercises (9 in all)
- Team exercise (Ex#1)……...2%
- Team exercises (Ex#2-Ex#3) ……..8%
- Individual Exercises (Ex#4 to Ex#9)..……....25%

Exams (3 in all; differentially weighted as noted below)..……..……45%
- Exam I: 15%
- Exam II: 30%
- Exam III: 55%

In Class Quizzes……..……..…10%

Class Participation & Attendance ……..……...10%

Total……..100%

*Bonus points for Presentation** 10%

Grading Notes:

Class Participation and Attendance. The following aspects will be considered in evaluating student's grade:

a. Preparedness and Attendance. Regular attendance and diligent preparation for the class is indicative of student commitment to participate in a learning environment. See attendance policy above.
b. Active Engagement. Active participation in the class with thoughtful questions, sharing alternative perspectives, and responding critically to class discussion, especially during case study discussions.

Bonus Points for Presentation. You will earn bonus points for making presentation to the class. Presentations will be graded as per policy for class assignments (see grade computation below). To earn bonus points, the students must earn 85% (B) or higher on their presentation. No bonus points would be earned for a presentation that earns less than a B. Presentation opportunity will be available as follows:

a. Invited Presentation. Students with outstanding submissions for homework exercises will be invited by the instructor to present their work to the class. Take this seriously. Imagine you are making a formal presentation to a client. Highlight your presentation skills.
b. Requested Presentation. Any students can request the instructor to make a presentation based on the class exercise. Such requests must be made at least 1 week in advance of the presentation date.

Exam. Each exam will test for comprehension of concepts and skills developed as part of the course, and their application to analytical problems.

Grade Computation: Most assignments will be graded using a plus and minus system (e.g., A- and A+ and so on). Conversion of this grading system to numerical scores is provided below. Students are responsible for tracking their own overall performance using the above weighting scheme. The final grade, however, will be a letter grade based on the final numerical score without the plus and minus in accord with University guidelines.

A+ = 98, A = 95, A– = 91
B+ = 88, B = 85, B– = 81
C+ = 78, C = 75, C– = 71
D+ = 68, D = 65, D– = 61
F+ = 50, F = 30, F– = 0

SUBMISSION REQUIREMENTS AND GUIDELINES.

All course exercises should be submitted ***only*** on the blackboard as per specific instructions to avoid late submission penalty (see below).

Typically, submission-feedback-rework will work as follows:

1. Assignments should be submitted in a powerpoint format, address all requirements of the assigned problem, and generally up to 10 slides of material. Text may be bulleted but must be complete with details to be understood precisely. Be concise. Avoid redundant words. Additional notes may be included in the powerpoint to explain your work.
2. Assignments will be due by Saturday, midnight. No extensions permitted (see ***Late Submissions*** below).
3. Feedback will be provided usually by Monday, 12 noon.
4. Based on the quality and creativity of your submission, the instructor may ask you to make a presentation based on your assignment in the Tuesday class with the following guidelines:
    a. *5-7 minute presentation*
    b. *Focus on points highlighted in the feedback. Revise and sharpen your presentation accordingly.*
    c. *Polished and professional presentation. Assume you are presenting to a client.*
5. Students who want to rework the submission to fix things that they missed or got wrong can submit their revised work by Tuesday, 12 noon (before class).
6. The main objective of the rework is to promote learning of key concepts and skills, and not rake up grade points (see ***Re-dos*** below).
7. All rework must be done individually by the student/group, without help, assistance or guidance from other students or course assistant.

Important*: Label all your assignment files as instructed below:*

All individual assignments should be labeled as your lastname_firstname_exercisename_MKMR 310

Example: singh_jagdip_ex3_MKMR310.pptx

All group assignments should be labeled as: last_names_of_group_members_exercise number_MKMR310

Example: singh_smith_thomas_payne_chin_Huella_MKMR 310.pptx

All reports should be carefully edited to eliminate redundant materials and keep the material focused. Tightly written papers will be viewed more favorably than long-winded developments.

Late Submissions: Late submission will result in a letter-grade penalty. That penalty is one full letter grade for each day (or part thereof) that the submission is late. For example, an exercise would have earned a B if submitted on its due date of Saturday, will be graded C if submitted by Sunday, D if submitted by Monday, and an F if submitted thereafter. If a submission must be late due to circumstances beyond your control, contact the instructor. At his discretion and based on his assessment of the actual degree of uncontrollability of the situation, he may permit a special arrangement. The most typical special arrangement is for students who must miss class due to extreme circumstances. They are often permitted to submit the assignment *early*. It is extremely rare for the instructor to permit an extension of the due date.

Redos. Students are encouraged to go back to fix problems identified on their graded assignments as this ensures that students complete the learning loop. In some instances, a redo may earn you a portion of the original points, but often it may not. Redos that show student's effort in addressing the problem on their own, the nature of the problem and timely submission are all factors that go into this decision. Additional points for redos are at the instructor's discretion. The extra credit will be upto 1 point maximum (10% of grade) and usually lower. Redos are intended to promote learning and not as opportunities for making up the grade.

Changes. The instructor reserves the right to make changes during the semester to any aspect of syllabus that, to his judgment, are needed to achieve the learning objectives of the course.

Weekly Schedule

***RR* = Required Reading, *WR* = Web-based material for Review,**
***HwA* = Homework Assignment**

Aug. 27: CLASS 1

Analytics Driven Insights: Basics

***WR**: Big Data and Analytics*, http://en.wikipedia.org/wiki/Big_data
http://www-01.ibm.com/software/info/rte/bdig/bdwa-7-post.html
http://thehumanfaceofbigdata.com/
http://en.wikipedia.org/wiki/Analytics
https://www.ibm.com/developerworks/community/blogs/techtrends/tags/analytics,2012?lang=en

Computer Lab: Basic Analysis with XCL (Univariate)

RR: MLGB: Chap 1

WR: "http://www.case.edu/its/training/lyndamain.html (LYNDA.com)
SUMIFs in Excel: http://www.youtube.com/watch?v=JonS_YuFeMQ
Pivot in Excel: http://www.youtube.com/watch?v=7zHLnUCtfUk
Vlookup in Excel: http://www.youtube.com/watch?v=981MeKFA4W8
Basic Statistics in Excel:
http://www.youtube.com/watch?v=efdRmGqCYBk&playnext=1&list=PL2C53784A0634D33B
Histogram in Excel 2010: http://www.youtube.com/watch?v=RyxPp22x9PU"

In-Class Activity: MLGB: IQs 1.3, 1.4 & 1.6

***HwA**: Ex#1: Due Aug. 31*

Sept. 3: CLASS 2

Dashboard Analytics with SPSS

***RR**:* SPSS Resources (PASW):
http://www.youtube.com/watch?v=5Z8RagPdmOI&feature=related

Computer Lab: Descriptive Analytics for Single Variables with EXCEL

RR: MLGB: Chap 3

***WR**:* "Define Variables and Marketing Terms
http://www.case.edu/its/training/lyndamain.html (LYNDA.com)
http://www.youtube.com/watch?v=ssjuOuMkCB4&feature=mfu_in_order&list=UL, &
http://www.youtube.com/watch?v=icgooM1K2lU&feature=mfu_in_order&list=UL
http://www.marketingpower.com/ResourceLibrary/Pages/common-marketing-terms.aspx"
Frequency Analysis in SPSS: http://www.youtube.com/watch?v=24Y5w9sAOzI
Selecting Cases in SPSS: http://www.youtube.com/watch?v=ero4VR7h1HU&feature=related

In-Class Activity: MLGB: IQs 3.2, 3.3 & 3.4;
Extra SPSS 3.4 & 3.5

HwA: Ex#2: "01.5 Data for Homework Movie_Sport_SPSS.sav" due Sept 7

Sept. 10: CLASS 3

Web Analytics & Monetizing

Real-world Problem Solving: Rosetta

***WR**:* Web Analytics 101: http://www.kaushik.net/avinash/web-analytics-101-definitions-goals-metrics-kpis-dimensions-targets/
Beginning Analytics: http://www.youtube.com/watch?v=Hdsb_uH2yPU&feature=relmfu
Advanced Analytics: http://www.kaushik.net/avinash/excellent-analytics-tip-11-measure-effectiveness-of-your-web-pages/

Computer Lab: Data Mining with XCL (Bivariate)

***WR**:* http://www.kaushik.net/avinash/google-analytics-is-re-launched-do-these-five-things-first-in-v2/

HwA: ***Ex#3 due Sept. 14***

Sept. 17: **CLASS 4**

Rosetta Case Study 1: Web Analytics in Practice
Attribution Modeling for Interactive Data

WR: http://www.kaushik.net/avinash/multi-channel-attribution-definitions-models/
http://www.kaushik.net/avinash/consumer-behavior-research-purchase-analysis/

HwA: **Ex#4 due Sept 21**

Sept. 24: **CLASS 5**

Rosetta Case Study 2: Attribution Modeling in Practice
Exam Review Session

Oct 1: In Class Exam I (material covered till 09/24)

OCT. 8: **CLASS 7**

Asking Questions to Gain Deep Insights

WR: Using Qualtrics for Mixed Survey Designs, http://www.qualtrics.com/university/market-surveys/
Survey design best practices:
http://www.wpi.edu/academics/ATC/Collaboratory/Idea/surveypractices.html
http://www.customersat.com/blog/best-practices-customer-survey-design

HwA: ***Ex#5 due Oct. 12***

Oct. 15: **CLASS 8**

Associational Analytics with SPSS

WR: "http://onlinestatbook.com/2/chi_square/one-way.html
http://onlinestatbook.com/2/chi_square/contingency.html"

Real-world Problem Solving: Cleveland Cavs

RR: MLGB: Chapters 5 & 8
WR: http://www.youtube.com/watch?v=p-ZaWgiM_Yg
In-Class Activity: MLGB: IQs 5.1, 5.2, 5.5. & 5.6; 8.2, 8.3, & 8.4
HwA: ***Ex#6 due Oct 19***

Oct. 22: **FALL BREAK**

Oct. 29: **CLASS 9**

Predictive Analytics with SPSS

RR: MLGB: Chap 9
WR: http://onlinestatbook.com/2/regression/intro.html
http://www.youtube.com/watch?v=4EFXic4sGdE
http://onlinestatbook.com/2/regression/multiple_regression.html
http://onlinestatbook.com/2/regression/influential.html"
In-Class Activity: MLGB: IQs 9.2, 9.3, 9.4 & 9.5
Exam Review Session
HwA: ***Ex#7 due Nov. 03***

Nov 5: In Class Exam II (material covered till 10/29)

Nov 12: CLASS 11

Computer Lab: Forecasting with XCL

WR: http://www.statsoft.com/textbook/time-series-analysis/?button=3#identifying (read sections on "Two General Aspects…" and "Trend Analysis."
http://www.kaushik.net/avinash/analytics-tip-calculate-ltv-customer-lifetime-value
http://www.youtube.com/watch?v=BHDblh9Y1-o&feature=related" sections on "Simple Exponential Smoothing," "Choosing the Best Value...," and "Indices of Lack of Fit."
http://analysights.wordpress.com/2010/05/06/forecast-friday-topic-moving-average-methods-2/ - moving average overview

HwA: Ex#8 due Nov. 16

Nov. 19: CLASS 13

Rosetta Case Study 3: Forecasting in Practice
A/B and Multivariate Testing

RR: MLGB: Chap 11
WR: "http://www.kaushik.net/avinash/experimentation-and-testing-a-primer/
http://www.kaushik.net/avinash/experiment-die-reasons-awesome-testing-ideas/"
http://onlinestatbook.com/2/analysis_of_variance/intro.html
http://onlinestatbook.com/2/analysis_of_variance/anova_designs.html
http://onlinestatbook.com/2/analysis_of_variance/one-way.html
http://onlinestatbook.com/2/analysis_of_variance/multiway.html"
http://www.kaushik.net/avinash/analytics-tip-calculate-ltv-customer-lifetime-value/

In-Class Activity: MLGB: IQs 11.2, 11.4, & 11.5

HwA: Ex#9 due Nov 23

Nov. 26: CLASS 15

Rosetta Case Study 4: A/B Testing in Practice
Exam Review Session

Dec. 03: In Class Exam III (Comprehensive material covered till 11/26)

Measuring Marketing Performance
MKMR 308

Weatherhead School of Management
Case Western Reserve University
(Spring 2014)

Instructor:	Rakesh Niraj
Email:	rakesh.niraj@case.edu
Class:	Tuesdays/Thursdays – 2.45 PM to 4.00 PM, meets in PBL 07
Office:	Location – PBL 235 Phone – 216 368 0799
Office hours:	Tuesdays/Thursdays at 4.15 PM and by appointment.
Online resource:	http://blackboard.case.edu

REQUIRED COURSE MATERIALS:

Text books (available at Case bookstore)

- Data Driven Marketing – by Mark Jeffery – Wiley – ISBN 978-0-470-50454-3
- I B M SPSS for Introductory Statistics (5th Ed.) – by Morgan, Leech, Gloeckner and Barrett – Routledge – ISBN 978-1-848-72982-7

Course Notes

Blackboard would contain some required and suggested readings as well as information related to the course such as announcements, grades, group projects and also other documents. Other relevant and topical material may be distributed in class.

COURSE DESCRIPTION AND OBJECTIVES:

Objectives

This course is aimed at training the next generation of marketing managers who are (1) well versed in data analysis and (2) are thoroughly familiar with metrics that measure marketing performance and are able to calculate as well as convey them to other managers, especially upper management, who are often interested in the financial results and metrics.

Skills

The course will have the following takeaways for the students:

- Familiarity with data management and manipulation as well as basic statistical analyses using IBM SPSS software program. It is expected that you will consider taking the IBM Analytics Certification after taking this class.
- Understanding and greater appreciation of links between marketing actions and financial outcomes.
- Ability to fully understand, calculate and convey a small number of important metrics that measure marketing metrics in various contexts.

Class Format

The class will be largely delivered in the computer lab and therefore, it will be intensely hands on. Note that even though all class meetings are in the computer lab, not all content will require computer work all the time. Many of you might have continued access to SPSS from your previous courses (like MKMR 310), but if you do not, then it is highly recommended that you buy SPSS student access through CWRU software center for your laptop computer. The SPSS part of the class will help you get familiar with (or reinforce your

familiarity with) the software and therefore, it would largely consist of discussions, problem solving and computer activities. The performance metrics part will also be hands on with the help of Excel spreadsheets, but may also have some more discussion on the underlying concepts and latest developments in the form of lectures and guest-lectures.

The tentative course content of the three modules of the class is given below:

(1) Data Management – with SPSS
   a. Reading Data from various sources
   b. Transforming and recoding data of various type
   c. Data cleaning – removing duplicates
   d. Aggregating data
   e. Merging files – by adding cases as well as variables
   f. Analyzing Multiple Response Questions
   g. Charts, tables and other outputs with SPSS

(2) Introductory Statistics – with SPSS
   a. Levels of Measurement
   b. Descriptive Statistics
   c. From sample to population – inferential statistics
   d. Comparing two groups – t tests and non-parametric tests
   e. Comparing multiple groups – ANOVA
   f. Association between two variables

(3) Metrics for conveying marketing performance
   a. Data driven marketing
   b. Essential non-financial metrics
   c. ROI: Profits, NPV, IRR and payback period
   d. Customer focused metrics
   e. Internet marketing metrics

CLASS ASSESSMENT:

Course grading will be done for a total of 500 points. Individual course-grading elements will not be curved and the final letter grade will be based on your absolute performance in all the aspects of the class. The table below summarizes the components you will be evaluated upon and the corresponding percentage distribution.

Individual Assignments

There will be a total of 5 individual assignments (two from the metrics portion and three from the SPSS portion) and the best 4 of these assignments will count (40 points each). These are individual assignments based on marketing data and examples. You need to complete them using appropriate software and then *create single Microsoft Word files for submission through Blackboard's SafeAssign* feature. Discussing the assignment with others is not only allowed, but encouraged. However, the final submission must be demonstrably your own work. Thus, it should not be copy of the same computer output as another student's submission, and the explanations and descriptions should be completely in your own words.

In-class computer based exercises

There will be in-class exercises almost every week. These will be short exercises of about 15-20 minute duration each. A total of 11 weekly exercises will be given and the best 9 will count (10 points each). These will usually be open-book, open notes, open blackboard material. External websites may be blocked for many of these.

Exams

There will be two in-class examinations (75 points each) in-lieu of mid-term exam for the SPSS part of the class (in weeks 5 and 9) and one final exam on the performance metric part (during exam week) will be given (100 points). The final exam will be non-cumulative. These will also usually be open-book, open notes, open blackboard material. External websites may be blocked for some of these.

Component	Percent of the Grade	Important Dates	Remarks
Individual Assignments (4)	32%	5 given – due in week 3, 6, 10, 12 and 14.	40 points each X 4 best
In-class exercises (9)	18 %	Almost weekly: except Weeks 4, 9 and 14 - Thursdays.	10 points each X 9 best
In-Class Exam I	15%	Week 4 - Thursday	75 points
In-Class Exam II	15%	Week 10 - Tuesday	75 points
Final examination	20 %	Exam week	100 points

Class attendance and participation

Since the classes are expected to be hands-on, fully engaged participation, including regular attendance is necessary. I will, however, not take attendance. There is no make-up test for the in-class exercises, except for documented illnesses and documented school-approved sports or other engagement. Since you can miss two of these without much adverse consequences, request for make-up tests will NOT be entertained except for the two reasons expressed above.

You will have a clear idea about what is to be covered in the class and you should be ready for the topic. There is cold calling in the class so students might be randomly picked as we go through the topics for the day. There will often a graduate assistant present to guide you through the hands-on portion of the SPSS software based instructions.

OTHER IMPORTANT ISSUES:

Teamwork

The class does not have a team-evaluation component. However, you are encouraged to work in teams. You are encouraged to make teams, meet me in teams for problem-solving and work on assignments together – of course, your final submission of the assignments should be strictly your own work and any writing should be in your own words.

Academic Integrity

Members of the class are expected to adhere to the principles of academic integrity that govern students at an institution of higher learning. For all the homework assignments, it is acceptable to have discussions with other classmates but the final submissions must be demonstrably your own work. You need to present the answer in your own words and do not turn in exactly the same computer output, for example. The two in-class examinations and the final examinations will be closed book and no help should be sought / given during these exams.

Special Needs

Please contact me as soon as possible with the required paperwork, if you need and qualify for special accommodations as per the university policy.

Instructor Feedback

Your instructor and the Weatherhead School are committed to continuous improvement in the quality of teaching and learning. Please feel free to speak to me (or email me) at any time about any aspect of this course, including things that you think are going well, or things that need to be improved.

Summary Weekly Schedule

Weeks 1 to 3: Data management and manipulation using IBM SPSS software [1 individual assignment given and 3 in-class exercise, every Thursday]

Week 4: Review and catch-up of Data management (Tuesday) and in-class exam – I (Thursday)

Weeks 5 to 8: Statistical analysis using IBM-SPSS software [2 individual assignment given and 4 in-class exercises, every Thursday]

Week 9: Review and catch-up of Introductory Statistics (Tuesday) and Start of 15 Metrics for Marketing (Thursday)

Week 10 (Tuesday): in-class exam–II {Moved to this week so there is some time after Spring Break}

Weeks 10 (Thursday) to week 14: The 15 metrics everyone in marketing should know [2 individual assignment given and 4 in-class exercises, every Thursday, except week 14] – Last class will be a review for final exam.

Final week: 2.5 hour final exam – May 1st, 12.30 PM to 3 PM.

Case Western Reserve University
The Weatherhead School of Management

Marketing Management

MKMR 301
Course Syllabus
Fall 2012

General Information

Instructor: Casey E. Newmeyer

Contact Information:
Email: cen12@case.edu
Office: 223 PBL
Phone: (216) 368 - 6404

Class Meeting Times: Tuesday and Thursday 1:15 pm – 2:30 pm
Class Location: PBL Room ???

Course Description

This course will provide students with a broad understanding of the marketing function in the modern world. It is designed to provide students both a theoretical and practical understanding of marketing decision-making. Topics covered in this course include the marketing environment for domestic and international marketing, marketing information systems, consumer and industrial buying behavior, target marketing, the formulation of marketing strategy involving the elements of product, distribution, promotion and pricing elements, and finally, market planning, organization, implementation, and control. Ethical considerations and a global view are incorporated into these topics, throughout the course.

The main learning objectives of this course include understanding:

- What marketing is
- How marketing is incorporated into the firm and its business plan
- The specific functions of marketing: product, place, price, promotion, people
- How marketing actions increase firm value

Required Materials:

- MKTG 5, 5th Edition
  by Lamb, Hair, McDaniel (ISBN-10: 1111528098 ISBN-13: 9781111528096)

In-class Assignments/Quizzes

During the course of the semester multiple quizzes and/or assignments will be given in class or announced and given in Blackboard. These could include material to be covered that day from the text. The point value for each quiz/assignment will vary based on the requirements of the quiz/assignment. If you are absent from class the day a quiz/assignment is presented you will receive a zero for that particular item. Make-ups are granted at the instructor's discretion and only for documented excuses.

Case Discussions/Questions

Two cases will be posted in Blackboard, for which you will prepare a typed response to open-ended questions. Remember to use proper MLA in-text citations when preparing your case responses and to use the proper format.

Class Participation and Discussion

Class participation will represent a substantial portion of the grade. Students will be expected to ask relevant questions, provide thoughtful comments and bring appropriate current-events examples of marketing and branding issues to class for discussion. Regular engagement is expected with the professor, fellow students and any guest speakers. The emphasis for grading will be on quality, not quantity.

You are expected to come to class thoroughly familiar with and ready to discuss assigned reading material and cases. Your simply showing up for class is useful for you, but adds little or no value to your colleagues' learning experiences. All students are expected to contribute to each others' learning. "Thoroughly familiar" means, not only being ready and able to discuss the main points addressed in the reading materials, but also being ready and able to discuss the nuances of their implications conceptually, for managerial decision making, and marketing performance.

Your class participation grade will also consist of unannounced in-class exercises and assignments.

Video Clip Assignment

This assignment requires the student to recognize course concepts in the mass media. This assignment consists of a brief class presentation and a one page summary presented to the instructor. The ideal assignment will identify one or two marketing principles depicted in a video or audio clip from a movie, TV program or in a radio segment. Students will work in teams of 2 and sign up for one of 4 presentation dates (Groups 1-4 on the course schedule). Further details will be posted on Courseweb.

Exams

Three regular exams and one final exam are scheduled for the semester. The final exam is cumulative. Exams could be a combination of multiple choice, true-false, matching, ordering, fill-in-the-blank, and short-answer.

Grades

- Video Clip Assignment 5%
- Class Participation 15%
- Case Write-ups (2 @ 7.5%) 15%
- Chapter Exams (3@10%) 30%
- Marketing Plan 10%
- Quant Project 10%
- Final Exam 15%

Course Schedule

Date	Topics	Reading	Assignment
August 28	Introduction and Syllabus Overview		Details of the deliverables needed will be discussed in-class and posted on blackboard
August 30	Chapter 1: An Overview of Marketing	Chapter 1	
September 4	Strategic Planning	Chapter 2	
September 6	Ethics and Social Responsibility and the Marketing Environment	Chapter 3 &4	
September 11	Developing a Global Vision	Chapter 5	
September 13	Case Discussion 1		Case 1 written questions due
September 18	Consumer Decision Making	Chapter 6	
September 20	Decision Making cont.		Group 1: Video Clip Assignment
September 25	Segmenting & Targeting Markets & Exam 1 Q&A	Chapter 8	
September 27	EXAM 1 – Chapter 1, 3, 4, 5, 6, 8		
October 2	Decision Support Systems & Marketing Research	Chapter 9	Group 2: Video Clip Assignment
October 4	Product Concepts	Chapter 10	
October 9	Developing & Managing Products	Chapter 11	
October 11	Products cont. Exam 2 Review		Group 3: Video Clip Assignment
October 16	Exam 2 – Chapters 9, 2, 10, 11 & 12		
October 18	Marketing Channels & Supply Chain Management	Chapter 13, 14	
October 23	Fall Break		
October 25	Integrated Marketing Communications	Chapter 16	
October 30	Advertising & Public Relations	Chapter 17	Group 4: Video Clip Assignment
November 1	EXAM 3 Chapters 13, 14, 19, 20, 16, 17 & 18		
November 6	Sales Promotion & Personal Selling	Chapter 18	
November 8	Thanksgiving Break		
November 13	Case 2 Discussion		Case 2 Written Questions Due
November 15	Pricing Concepts	Chapter 19	
November 20	Setting the Right Price	Chapter 20	
November 22			Quant Project
November 27	Group Presentations		Marketing Plan Presentations
November 29	Group Presentations		Marketing Plan Presentations
December 4	Social Media	Chapter 22	
December 6	Course Wrap-up/Final Review		
December 13, 12:30pm	FINAL EXAM		

CWRU Academic Integrity Policy
All students in this course are expected to adhere to university standards of academic integrity located at: http://studentaffairs.case.edu/groups/aiboard/policy.html. You are responsible for familiarizing yourself with the University Honor Policy. All students in this course are expected to adhere to university standards of academic integrity. Cheating, plagiarism, and other forms of academic dishonesty will not be tolerated in this course. This includes, but is not limited to, consulting with another person during an exam, turning in written work that was prepared by someone other than you, and making minor modifications to the work of someone else and turning it in as your own. Ignorance will not be permitted as an excuse. If you are not sure whether something you plan to submit would be considered either cheating or plagiarism, please do not hesitate to ask me. Any cheating, plagiarism, or other forms of academic dishonesty will result in a grade of "F" for this course, as well as the notification of the appropriate Weatherhead School of Management authority.

Disability
Students at Case Western Reserve University are not required to disclose disability information to anyone. However, in order to use services and appropriate accommodations, students should notify Disability Resources in Educational Services for Students at 216.368.5230 or disability@case.edu. Disability Resources is located in Sears room 470.

Basic Statistics for Engineering and Science

Stat 312 – Spring 2014

Syllabus

Instructor: Danhong Song, Ph.D
Room 375, Yost Hall 216-368-2918 danhong.song@case.edu

Office hours: 2:40 – 4:00 pm, Tuesdays and Thursdays, or by appointments

Class Time: T R 1:15 – 2:30 pm

Location: Room 400, Nord Hall

Prerequisites: MATH 122 or equivalent.

Text Book: Engineering Statistics, 5th Ed, Montgomery, Runger, and Hubele, Wiley, 2011, ISBN-13 978-0-470-63147-8

Coverage: Chapters 1 – 6 (tentatively):

Ch 1 The Role of Statistics in Engineering
Ch 2 Data Summary and Presentation
Ch 3 Random Variables and Probability Distributions
Ch 4 Decision Making for a Single Sample
Ch 5 Decision Making for Two Samples
Ch 6 Building Empirical Models

Homework: Homework problems will be assigned regularly along with a due date (usually Thursdays). Homework must be turned in, on the due date, in class, before the class starts.

Exams: There will be two in-class exams (75 min.) during the semester and an in-class comprehensive final exam (3 hours) at the end of the semester on the following dates. Students must bring a calculator. However, cell phones are not permitted to be used as calculators for the exams. Make up exams may be given with an advanced notice.

Mid-terms: Tuesdays, Feb. 18 and April 8 (tentatively)
Final Exam: 12:30 - 3:30 pm, Monday, May 5, 2014.

Grading: The final grade will be based on a composite score of all the points coming from the exams and homework. The break up for the final grade will be as follows:

Homework:	20%
Mid-Terms:	45%
Final Exam:	35%

A: 90 -100 B: 80 - 89 C: 70 - 79 D: 60 - 69 F: 59 and below

Website: The course website is http://blackboard.case.edu.To log in you need to use your case ID and password. After logging in, go to the appropriate **STAT 312** under "My Courses".

Help: Hayley Topel (hmt17@case.edu) and Janis Cava (jac256@case.edu) will be my TAs for this course during the semester. They will provide tutoring help and share the responsibilities of grading. For tutoring, both Hayley and Janis will be able to tutor on Mondays in room 374 Yost. Here are their hours:

Hayley: 10:00-11:30 am on Mondays and
Janis: 4:40-5:30 on Mondays

Also, please don't forget about my office hours. I will be glad to help you.

NOTE: The instructor reserves the right to make any changes she considers academically advisable. It is the student's responsibility to attend classes and keep track of the proceedings.

STAT 201
Spring 2014
TR 1:15-2:30 PM, Wickenden 301

Instructor: *Dr. Patricia Pepple Williamson*
342 Yost Hall, 368-6013
email: patricia.williamson@case.edu
Office Hours: 1:30-3:30 MW, 12:00-1:00 TR

Textbook: *Introduction to the Practice of Statistics, 7th edition, by Moore McCabe, and Craig.*

Grading: *Two tests and a comprehensive final exam will be given. Fifteen take home quizzes will be given of which the top 12 will be counted. You are not to work with anyone on the quizzes; each quiz should be your own work. If there is a question on a quiz, consult me.*

Quizzes	*- 120 points (25%)*
2 Tests	*- 200 points (41.7%)*
Comprehensive Final	*- 160 points (33.3%)*
Total	*- 480 points*

Grading Scale:

90 - 100%	*A*
80 - 89.99%	*B*
70 - 79.99%	*C*
60 - 69.99%	*D*
59.99% and below	*F*

Make-up Policy: *No make-up tests or quizzes will be given. Students who know in advance that they must miss a test due to exceptional circumstances may be able to take the test in advance. Please inform me of any problems as soon as possible.*

Help: *Please feel free to see me about any questions you have during my office hours; other times can possibly be arranged. I will be glad to help you.*

Withdrawal: *Last day to withdraw with a mark of "W" is March 28, 2014 (April 28 for first year undergraduates).*

Student Conduct: *Be advised that students are expected to behave appropriately in class. Please turn off cell phones as well.*

Attendance: *Attendance is important for adequate student comprehension. Also, to have 15 take home quizzes, it is obvious that they are given out frequently so you need to be in class to get the quizzes. If you have to miss a class, it is your responsibility to find out what you missed. Students must be in class to obtain a quiz except at most two quizzes that can be sent via email with acceptable excuses.*

Homework: *Problems will be assigned regularly throughout the course, but will not be collected or graded. Keeping up with homework assignments is the responsibility of the student and plays an important part in the mastery of course material. Answers to most odd problems are available in the back of the textbook.*

Website: *The course website is http://blackboard.case.edu.To log in you need to use your caseID and password. After logging in, go to* ***STAT 201*** *under "My Courses". I strongly suggest you copy the Power Point slides for each chapter/section (posted under "Course Documents") and bring the appropriate chapter notes to class.*

Calculator Use: *Cell phones are not permitted to be used as calculators for tests.*

Computer Access: *I expect students to have access to a computer where they have Excel, Minitab, and R; R can be downloaded from the internet. The instructions to download R are given below.*

To download R:

1. *Go to www.google.com.*
2. *Type in "R".*
3. *Click top entry.*
4. *Click on CRAN mirror.*
5. *Under USA, click on http://cran.case.edu/*
6. *Click on Windows for those with windows.*
7. *Click on base.*
8. *Click on Download R 3.0.2 for Windows.*
9. *Continue through process as directed.*

You should end up with R on your desktop.

Of course, if you don't have windows, adapt accordingly.

EECS 132 Spring 2014

DESCRIPTION

Course Description

Introduction to computer programming and problem solving with the Java language. Computers, operating systems, and Java applications; software development; conditional statements; loops; methods; arrays; classes and objects; object-oriented design; unit testing; strings and text I/O; inheritance and polymorphism; GUI components; application testing; abstract classes and interfaces; exception handling; files and streams; GUI event handling; generics; collections; threads; comparison of Java to C, C++, and C#.

Course Outline

The course will cover the following topics. The order is subject to change:

- A brief history of programming
- Data, primitive data types, class data types
- Methods
- Computer organization and the Java memory model
- Conditional statements, recursion
- Loops
- Strings and Arrays
- Polymorphism, subclasses, abstract classes, interfaces
- Industrial programming techniques: paired programming, API generation
- Graphical User Interfaces and event-driven programming
- Introduction to computer science: program correctness and runtime analysis
- Industrial programming techniques: testing, debuggers, and JUnit
- File input and output
- Building applications; the main method
- Exception handling
- Abstract data types; generic types; linked lists
- Collections; iterators; comparable types
- Inner and anonymous classes
- Java reflection
- Multi-threaded programming

COURSE INFORMATION

Instructor: Harold Connamacher

- Office: Olin 502
- Office Hours: Monday 4-5pm, Tuesday 9-10am, Wednesday 4-5pm, Thursday 9-10

Timetable:

- Lecture: Monday, Wednesday, Friday, 2:00pm in Millis Schmitt Lecture Hall

- Recitation Lab: Wednesday 3:00 pm **-or-** Wednesday 4:00 pm **-or-** Thursday 1:15pm **-or-** Thursday 2:45p in Olin 803

Because all recitation sections are full, please attend your assigned recitation lab. If you need to change recitation sections, please speak to the course instructor.

Course Textbook:

- **Required Book:** Flanagan, Java in a Nutshell, 5th edition, 2005. ISBN 9780596007737
- **Recommended Book:** Lewis and Lofton, Java Software Solutions, 7th edition, 2012 (with electronic resource access). ISBN 9780132760775

The second book is strongly recommended for any student who does not have prior experience with computer programming.

Prerequisites:

- None

COURSE ASSESSMENT

To help you learn the material, there will be regular assignments and assessments covering different aspects of programming in Java.

Late Work Policy: The late policy for each assessment is listed in the assessment description. The only exceptions are for medical or similar emergencies, and in such situations work will only be accepted with notice from an appropriate university official. If you have to miss class due to a scheduled event such as a sporting event or club performance, you must make arrangements to submit your required work early.

Prelabs: The prelab is a short exercise to be completed each weekend. The prelab is due each Wednesday at 2pm beginning Wednesday, January 22. **No late submissions will be accepted.** The purpose of the prelab is to reinforce certain Java techniques that will be used in the next lab session so that your lab session will be more productive. The prelab is expected to be entirely your own work, and you will receive full credit as long as it is mostly correct.

Quizzes:There will be a short quiz to be completed each week. The quizzes are posted each Friday, starting January 17 and due the following Tuesday. **No late submissions will be accepted.** The quiz will cover material from the previous week lecture. The quiz is expected to be your own work. You will be permitted to take the quiz twice. After the first attempt, you will be told which answers were incorrect and you may look up the correct answers in your notes or the textbook.

Optional MyProgrammingLab:MyProgrammingLab is an on-line Java quiz/program practice that goes with the Lewis textbook. MyProgrammingLab is optional and recommended for students with no prior programming experience. New MyProgrammingLab questions will be posted each week. If you choose to do the MyProgrammingLab, you may replace either your term Prelab or Quiz grade with the term MyProgrammingLab grade

MyProgrammingLab access is included when you purchase the Lewis textbook. **You may purchase MyProgrammingLab separately.** Go to myprogramminglab.com to register.

Labs:There will be a lab/recitation section every week beginning Wednesday, January 15. The lab will consist of an exercise you are to complete in the hour provided. The lab serves two purposes: to give you hands on practice with Java concepts, and to introduce you to the industrial technique

of paired-programming. You are encouraged to have a lab partner. If you have a lab partner, your lab grade will be determined by how well you follow the paired-programming technique and whether you focused on the lab task for the entire hour. If you do not have a lab partner (for example you miss the lab due to illness), you will not be practicing paired-programming and so your grade will be determined only by the correctness of your lab results. **If you are unable to attend the lab session, you can have an automatic extension. The lab must be submitted no later than the Tuesday following the lab section.**

Programming Projects: There will be a programming project assigned every two weeks starting the fourth week of the course. The first project will be assigned Friday, February 7 and will be due Friday, February 21. Each homework will consist of Java programming. In addition, there will be written work to include with the program. All homework is assumed to be your own work. **The late policy for homework is as follows: -10% if within one hour late. -25% if within 24 hours late. -50% if over 24 hours late. There will be no homework accepted once the grading is done or after a week past the deadline.**

Midterm:The midterm exam will consist of short answer, fill in the blank, and free response sections. You may be asked to provide correct Java code or proper English descriptions for the questions asked. The midterm will be in class on Monday, March 3.

Final: The final exam is at the time specified by the registrar: Monday, May 5 at 8:30 am. The exam will be the same format as the midterm exam, and the exam will be cumulative.

Calculators: You are permitted to use calculators on prelabs, labs, homework, and quizzes. No calculators will be permitted on midterms and finals.

Grading Scheme:

Prelabs: 5% *
Labs: 10%
Quizzes: 5% *
Homework: 30%
Midterm: 15%
Final Exam: 35%
* You may replace either the 5% quiz grade or the 5% prelabs grade with a 5% MyProgrammingLab grade.

Grade scale:
The class will use the following scale to translate your percentage grade to a final grade.

If your percentage grade is at least	Your final grade will be at least
90	A
80	B
70	C
60	D

You must score at least 50% on the final exam to pass the course.

Academic Honesty

Please see the general University Policy on Academic Integrity. The specifics for the course are

listed in the document, Course Honor Policy, found on this site.

Special Considerations

Physical Disabilities or Other Hardships: If you have a physical disability or other hardship that can potentially put you at a disadvantage in this course, please see Educational Student Services. They will make certain you receive the necessary accommodations so that you may perform your best.

Religious Holidays: I strive to schedule all major projects and tests so that they do not conflict with important religious holidays. However, I am not always successful in doing that. If an important religious holiday conflicts with a class test or assignment in a way that makes it so that you can not take the test or complete the assignment as originally assigned, please see me as soon as possible to make necessary arrangements.

ENGR 131: Elementary Computer Programming

Instructor: Dr. Chris Fietkiewicz
Email: Use help addresses listed on the website
Instructor Office Hours: Immediately after lectures and by appointment
Teaching Assistant Office Hours: See the website

Websites

Primary website: http://blackboard.case.edu
Textbook source code: http://www.elsevierdirect.com/9780123850812 (click on link for "Companion Website")

Getting Help

Email: Use the "help" email address for your lab section (see the website for the list). Email sent to this address goes to several TAs for the fastest reply possible.

Office hours: Supplemental Instruction (SI) sessions will be available on different days throughout the week. Each TA will also hold weekly office hours. See the website for the complete schedule of SI sessions and TA office hours.

Tutoring: Complimentary tutoring is available through ESS Peer Tutoring services. To sign up, go to the ESS web page, http://studentaffairs.case.edu/education/, and click on "TutorTrac". Tell your tutor to contact Chris if he or she has any questions.

Course Description

The goals of this course are to (1) develop skills in computational thinking through problem solving and (2) learn the practical skill of programming in MATLAB. Topics covered include algorithm design, data structures, operators, control flow, and functions.

Textbook

Book: Attaway, *MATLAB: A Practical Introduction to Programming and Problem Solving*, Second Edition, Butterworth-Heinemann, 2011. ISBN 978-0123850812.

Website with M-files: *see link above*

Specific Topics

Ch 1: MATLAB calculations
Ch 2: MATLAB programming
Ch 3: Selection
Ch 4: Repetition
Ch 5: Vectorization
Ch 6: Advanced programming
Ch 7: Strings
Ch 8: Data Structures

Ch 9: File Input/Output
Ch 10: Advanced functions
Ch 11: Advanced plotting
Ch 12: Systems of linear algebraic equations
Ch 13: Statistics and advanced algorithms
Ch 14: Audio and graphics
Ch 15: Advanced mathematics

Labs & Software

All students will have 24-hour access to the Olin 8th floor lab (x4056). The Nord Lab computers can also be used. The Olin lab and all equipment are to be used for coursework only. Any unauthorized use of the lab or equipment may result in failure for the course. MATLAB software is available from https://softwarecenter.case.edu, and students are encouraged to install in on their personal computer.

NOTE: To use MATLAB off campus, you may need to use the AnyConnect software for virtual private networking (VPN) which is available at https://vpnsetup.case.edu.

Attendance

Lab: Lab attendance is required, and students must be present to receive credit for lab assignments. If you have a valid excuse for missing a lab meeting, contact your TA immediately using the "help" email address.

Lecture: Attendance is not required for actual lectures (including lectures after a quiz). However, lecture activities will be collected randomly to check attendance which can improve your course grade (see *Borderline Grades* below). Students with valid excuses for missing a lecture can request makeup work in order to obtain attendance credit.

Respect: Except for excused absences, it will be assumed that students who are not present during lecture or lab do not need further assistance with the material covered. Please be respectful and use our scheduled time on Tuesdays and Thursdays to seek help. At the end of the semester, grades within 1 point of the next highest letter grade *might* be raised at the instructor's discretion for students with perfect attendance and perfect assignment completion.

Grading

The grade cutoffs are 90-80-70-60 percent for A-B-C-D respectively. The following weights will be applied to the average percentage for each category. Items from different categories may have the same total number of points, but they do not have the same contribution to the final grade. For example, an exam and a quiz may each have a maximum of 100 points, but the contribution of each depends on the number of exams, the number of quizzes, and the weight of each category.

(15%) Lab Assignments: Lab assignments are collected every week in lab. Students must be present to receive credit. Assignments cannot be made up except for excused absences.

(5%) Homework: Homework problems will graded for completeness and not correctness.

(20%) Lecture quizzes: Quizzes will occur at the beginning of selected lectures. They will cover material in the homework problems.

(50%) Exams: There are three required midterm exams. An unexcused absence will result in a zero for that exam. Students will have the option (not required) to take a comprehensive final exam that will count as a fourth exam. The average of all exams taken (either 3 or 4) is worth 50% of the final grade.

(10%) Final Project: A software development project is assigned at the end of the semester. Students must choose a particular design within the guidelines provided. Students are required to work in groups of two or three. Students are not permitted to work independently.

Late submissions: Late submissions for lab assignments and final projects will receive a penalty of –0.07 points for each minute past the due date. Exams and quizzes will not be accepted late except for excused absences (see "Abscences" below.

Absences: Non-emergency absences must be approved in advance, including Case-sponsored trips. Makeup exams will be only be provided for excused absences in accordance with university policy. Students with valid excuses for missing a lecture can request makeup work in order to obtain attendance credit.

Borderline grades: At the end of the semester, students who are within 1% of the next highest letter grade may receive the higher letter grade, at the instructor's discretion, if all of the following conditions are met:

- The student has perfect lecture attendance (based on collected lecture activities).
- The student submitted all assignments, including every lab, homework, quiz, exam, and final project component.
- The student received a score of 80 or higher on the final exam.
- The student has not violated any course policies.

Assignment Philosophy

There are different types of assignments in the course:

- Tutorials (*not* graded)
- Lab assignments (graded for correctness)
- Homework problems (graded for completeness, *not* correctness)

These are specifically designed to work together to prepare you for the quizzes and exams. Each one requires an additional level of ability. A homework problem is the most important opportunity for a student to practice with a large problem independently prior to a lecture quiz. Therefore students are encouraged to complete all homework problems independently. Asking for help is fine, but your results will be superior if all work is your own.

Academic Integrity

All homework submitted should be the student's own work. This includes the design of programs as well as the actual computer code. Students are encouraged to seek assistance from their TA and the instructor to ensure that homework study is approached properly. Students found in violation of the university's academic integrity policy will be reported to the Dean's office and will receive either a permanent zero on the assignment in question or failure for the course.

Friendly Advice: The Proactive Student

The truly successful people in life are the ones who know how to get what they want. They also know they are the only ones responsible for making it happen. Many students come to college thinking they can succeed by simply doing what they are told. It is not uncommon to hear the complaint, "You didn't tell us we had to know *that*!" Success in college, as well as in life, comes from *self*-discovery, which includes discovering things *by* oneself as well as discovering things *about* oneself. One of the instructor's jobs is to show students where to look. It is the students' responsibility to know whether they have found what they needed and to seek out help when they are uncertain. Please be proactive about your education, and do not wait for your instructors to tell you when you don't know something.