UNC Chapel Hill Faculty Working Group on Data Studies Curriculum

Developing Data-Literate Students

Acquiring, Managing, Analyzing, and Using Data in Societal Context

Report and Recommendations

April, 2014

THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

UNC Chapel Hill Faculty Working Group on Data Studies Curriculum

Developing Data-Literate Students

Acquiring, Managing, Analyzing, and Using Data in Societal Context

Report and Recommendations

April, 2014

TABLE OF CONTENTS

Page

SUMMARY 1
BACKGROUND 3
GOAL FOR UNC CHAPEL HILL 4
Target Audiences 4
ASPECTS OF DATA STUDIES 5
Definitions: Data Studies and Data Literacy 5
Ubiquity of Data 6
Societal Segments 7
Social, Ethical and Legal Challenges 7
Research Practices and Outcomes 8
Impact on Industry and Organizations 10
UNC RESPONSE 13
Support from the University Libraries 14
Curriculum Experts 15
Share These Findings 15
Ask the Provost to Form a Steering Committee 15
Create Awareness 16
Build a Curricular Resource Center 16
BUDGET 17
CONLUSIONS: THE ROAD AHEAD 17
APPENDICES 19
Appendix A: Faculty Working Group Member List 19
Appendix B: Kenan Science Library Report of Data Science Education, Student: Erin Morris 23
Appendix C: UNC Data Studies Course Inventory 47
Appendix D: Kenan-Flagler Business School Proposed Teaching and Other Opportunities in Big Data, Faculty: Noel Greiss 87
Appendix E: Kenan-Flagler Business School Dean's Fellows Project: Curricular Opportunities in Big Data, Students: Lara Koch, Jon Holbrook, Chris Huchenski, Parker McAllister 90
Appendix F: Computational Social Sciences for Undergraduate and Graduate Students, Student: Andrew Cabaniss 125

THE UNIVERSITY
of NORTH CAROLINA
at CHAPEL HILL

UNC Chapel Hill Faculty Working Group on Data Studies Curriculum

Developing Data-Literate Students

Acquiring, Managing, Analyzing, and Using Data in Societal Context

Report and Recommendations

April, 2014

Faculty Working Group Co-Chairs: Richard Marciano, Judith Cone. Committee: Stan Ahalt, Robert Allen, Daniel Anderson, Ray Angle, Nilay Tank Argon, Sridhar Balasubramanian, Stephanie Barnwell, Michelle Bolas, Peter Brews, Neal Caren, Thomas Carsey, Robin Cunningham, Terry Furey, Jan Hannig, Carol Hunter, Paul Jones, Kip Kelly, David Knowles, Ashok Krishnamurthy, Pam Lach, Anselmo Lastra, James Leloudis, Peter Leousis, Shannon McKeen, Danianne Mizzy, Peter Mucha, Ram Neta, Andrew Nobel, Vladas Pipiras, Arcot Rajasekar, Rachel Seidman, Ryan Thornburg, Barbara Wildemuth, with assistance from Kennetha Smith.

> *The explosion of data and the emergence of computational data analysis as the key scientific and economic approach in contemporary societies create new kinds of divisions. Specifically, people and organizations are divided into three categories: those who create data (both consciously and by leaving digital footprints), those who have the means to collect it, and those who have expertise to analyze it. The first group includes pretty much everybody in the world who is using the web and/or mobile phones; the second group is smaller; and the third group is much smaller still. We can refer to these three groups as the new "data-classes" of our big data society.*[1]

SUMMARY

This report presents the work of the UNC Chapel Hill Faculty Working Group on Data Studies Curriculum (Appendix A). The members of the Working Group believe the following:

[1] L. Manovich, "Trending: The Promises and the Challenges of Big Social Data", Debates in the Digital Humanities, ed M.K. Gold. The University of Minnesota Press, Minneapolis, MN, 2011 (http://www.manovich.net/DOCS/Manovich_trending_paper.pdf)

> Data literacy has become essential to research and scholarship, to learning at all levels, to translational endeavors, and to future student career success. In fields from healthcare to the humanities; the acquisition, management, analysis, and use of data has become a required skillset for college graduates. Further, because Carolina is committed to having its faculty and students see the world broadly and think critically and multi-dimensionally, data literacy should be embedded in an understanding of the influence of data on individuals and society. This contextualization of data includes examining such areas as the effects of data proliferation on social constructs, communication, privacy, security, and ethical considerations.

The Working Group chose the term data **studies** rather than data **sciences** in order to reflect this broader approach. Data **science** is a term more associated with data analytics and computer science. The Faculty Working Group believes that a degree from UNC Chapel Hill should give students a competitive advantage in their ability to work with data, regardless of a student's field of study or level of technical expertise in computer science.

Data Studies = Technical Data Skills (acquire, manage, analyze, use) in Societal Context

The Working Group recommends that UNC Chapel Hill offer a broad range of data studies curricular options for its students. This may be accomplished at the graduate and undergraduate level through course modifications, new courses, concentrations, minors, majors and certificate programs. The competencies taught will vary according to field of study, sub-specialties, the requirements of particular disciplines, and whether students are undergraduates or graduate students.

Within the broader scope of data **studies**, some academic units will delve into the technical aspects of data **science** more deeply**,** while others will spend more time on the social science components. The Working Group hopes a holistic approach of including societal implications, even in the most technical of courses, will emerge. For instance, the business school is planning courses in data analytics and is recognizing the relevance of privacy and ethical discussions in this technical subject. The graduate certificate program in digital humanities (open to all graduate students), launched in the Fall of 2013 by the Carolina Digital Humanities Initiative, focuses on humanities topics while also teaching technical competence in acquiring, managing, analyzing, and using data. The ultimate aim of this Working Group is for UNC students to be data literate in both the technical and social science aspects, with the balance determined by the type and level of student and the discipline.

BACKGROUND

This focus on data studies emerged as part of UNC Chapel Hill's Innovate@Carolina work, which is a strategic initiative to more effectively put university-born ideas to use for the public good. It is only natural for Carolina, as the first public university, to have a renewed focus on translation. The UNC Chapel Hill Board of Trustees and senior administrators are committed to advancing innovation and impact, and are focused on the fundamental research and teaching steps needed to achieve this[2]. They want to ensure that Carolina is a place where innovators thrive.

After studying the reasons Carolina may or may not be prepared for this translational work, it became apparent that UNC had to develop and/or strengthen some fundamental areas. Applied physical sciences, biomedical engineering, and computer science were identified as important to an innovation agenda for the campus. Also highlighted was the importance of working in multidisciplinary teams to address major issues, especially bringing the humanities together with the hard sciences and professional schools. To advance its capacity for working across disciplines, UNC adopted its first campus-wide theme: *Water in our World*[3]. The success of this pan-campus initiative is evidence of the potential for cross-discipline collaboration and the need for students and faculty to continue developing capacity to work in diverse groups.

Amid these conversations, data literacy emerged as a key component necessary to enable student innovation and equip students to compete today and in the future. In response, a group of 35 faculty and staff from 15 disciplines and areas convened to explore the subject of undergraduate and graduate curricula in data studies. The process was open and inclusive, with input sought from additional departments and curriculum experts on and off campus. The group adopted three underlying principles for its deliberations: urgency, inclusiveness, and interdisciplinary collaboration.

The Working Group quickly learned that other universities are moving rapidly to prepare their students. A survey of graduate and undergraduate offerings from universities compiled by UNC Libraries in 2013 (Appendix B) shows this trend. This excellent report by Erin Morris of the Kenan Science Library on the current state of post-secondary education programs in data science and related disciplines in U.S. universities discusses the impetus for such programs, curriculum trends, and requirements for a number of programs at both the graduate and undergraduate levels.

[2] http://innovate.unc.edu/site/wp-content/uploads/2012/10/Innovate_at_Carolina_Roadmap.pdf
[3] https://watertheme.unc.edu/

Many universities already have established degree or certificate programs in data science, analytics, or data management. Based on our preliminary research of other university programs, the Working Group believes that if UNC were to make data literacy an integral part of the curriculum, it would be the first major research university to adopt data studies using a holistic approach.

GOAL FOR UNC CHAPEL HILL

Prepare graduate and undergraduate **students to be data literate** in order to increase their competencies for academic pursuits, raise their competitive advantage in the marketplace of ideas and jobs, and increase their translational skills. This data skillset, to be customized with appropriate technical depth and embedded in a contextual understanding of the broad societal implications, includes:

- Contextual understanding – societal implications
- Critical and analytical thinking
- Problem solving
- Data analysis (drawing conclusions)
- Communicating and presenting findings
- Decision making
- Basic technical knowledge related to databases, code, storage, and security

The Working Group wants data studies to be available to all students in some form. The group noted that there seems to be a natural separation between students who have strong proclivities toward science/math and those who do not, at both the undergraduate and graduate levels. For example, while a course such as data analytics might be offered to all students, the curriculum could be developed with more rigorous technical requirements for science/math-intensive students, and tailored differently for their counterparts (general). The four main subsets of students requiring different approaches to content development are:

- Undergraduate — general
- Undergraduate — science/math-intensive
- Graduate — general
- Graduate — science/math-intensive

The Working Group has created a **repository of resources** to help faculty as they develop programs, courses, and activities for preparing undergraduate and graduate students with fundamental knowledge and skills in data studies. During the course of researching this topic, a

Sakai site served as the home for research materials. This content has now been moved to a website (https://datastudies.web.unc.edu) and can serve as a first stop for information on current courses, reference people on campus, and materials from other universities and industry. An open question for the next phase will be whether or not to establish campus-wide learning outcomes and metrics for the skillsets mentioned above in each of the student subsets or to leave learning outcomes to each department or school.

ASPECTS OF DATA STUDIES

Definition

The term *data studies* was adopted by the Working Group as more encompassing than the commonly-used term *data sciences*, which implies a specific set of content that is technical in nature. The Working Group offers this definition:

> Data studies combines data acquisition, management, analysis, and use of data (including a level of technical competence appropriate for each student) with an understanding of the nature of data and its broader implications for society.

While the Faculty Working Group would enhance the following definition of data literacy provided in the *Data Journalism Handbook* to include the contextualization of data in society, it found it helpful:

> Just as literacy refers to "the ability to read for knowledge, write coherently and think critically about printed material" data literacy is the ability to consume for knowledge, produce coherently and think critically about data. Data literacy includes statistical literacy but also understanding how to work with large data sets, how they were produced, how to connect various data sets and how to interpret them.[4]

Supporting the proposed more holistic approach to teaching data studies, there is evidence that contextual as well as technical competencies are becoming recognized as critical components of data literacy. In its second round of Big Data Funding, the White House recently launched a "Council for Big Data, Ethics, and Society."[5] The focus of this Council is to understand the social, ethical, legal and policy issues that underpin the big data explosion.

[4] http://datajournalismhandbook.org/1.0/en/understanding_data_0.html#sthash.I4cVpUnE.dpuf
[5] http://www.datasociety.net/initiatives/council-for-big-data-ethics-and-society/

The range of topics covered by data studies includes acquiring, managing, analyzing, and using data. All of this happens in an important iterative lifecycle associated with this process. When the lifecycle is not understood, the value derived from data is a) limited in scope, b) sub-optimized in terms of actual results, and c) lacks the scale that organizations need to drive sustainable improvements.

LIFECYCLE MANAGEMENT

ACQUIRE	MANAGE	ANALYZE	USE
Seek and discern available data. Collect, assemble all types of data – structured and unstructured: Voice, text, log files, images or video. Known and unknown data.	Store, secure, archive, administer data platforms, projects, and tools. Metadata, auditing, semantic knowledge representations, curation.	Inspect, clean, model, mine, visualize.	Ascertain what the data is revealing. Translate into actionable recommendations. Apply.
SOCIETAL CONTEXT – Complexity, multi-disciplinary, social structures, socialization, brain science and learning, security, privacy, ethics, policies and social norms.			

Ubiquity of Data

The availability and integration of data is changing and often disrupting disciplines, organizations, and industries across sectors. Yet data alone are useless. To be of value, data must be structured and analyzed in a timely manner for insights that can inform decisions and improve performance. Erik Brynjolfsson, Professor of Information Systems at Massachusetts Institute of Technology, suggests that "the big problem is going to be the ability of humans to use, analyze and make sense of the data."[6] Likewise, Jason Burke, Senior Advisor, Innovation and Advanced Analytics, UNC School of Medicine & UNC Health Care System states in his book, *Health Analytics: Gaining Insights to Transform Healthcare,* 2013, "complexity exceeds cognition." He goes on to say that our ability to understand the nature of the world around us will increasingly require us to be more creative in how we aggregate and analyze data.

[6] Steve Lohr, "For Today's Graduate, Just One Word: Statistics," The New York Times, Aug. 5, 2009

Societal Segments

The driving need for improving data literacy can be seen in three significant areas:

- Social, ethical and legal challenges
- Research practices and outcomes
- Industry and organizations, expectations of employers

Social, Ethical and Legal Challenges

Newborns today arrive in the world with a digital footprint already firmly established. Expectant parents post ultrasounds on Facebook and share moments of a child's early years with family and friends through social networking. The amount of personal information about us that is tracked online (or offline) increases daily. An emerging health care aspect is how consumer data – which is not covered under the United States' health privacy regulations like HIPAA – can be used to infer health issues such as an individual's risk of disease. For instance, if a health insurer offers a consumer a different insurance product based on Harris Teeter VIC card data, are we as a society accepting of that?

In this fast-paced and high-tech environment, massive amounts of personal raw data are available for mining by organizations and individuals. The need for an ethical framework around accessing, using, and sharing this information is critical. MIT has recently announced a "Big Data Working Group on Privacy." The goal is "to think long term to better understand and help define the role of technology in protecting and managing privacy, in particular when large and diverse data sets are collected and combined."[7]

The shifting social landscape also requires an understanding of the emerging digital divides created by the explosion of data, the implications across economic and societal sectors, and the challenges of addressing these issues for the future. Manovich, in his book, *Trending: The Promises and the Challenges of Big Social Data,* states:

> The explosion of data and the emergence of computational data analysis as the key scientific and economic approach in contemporary societies create new kinds of divisions. Specifically, people and organizations are divided into three categories: those who create data (both consciously and by leaving digital footprints), those who have the means to collect it, and those who have expertise to analyze it. The first group includes pretty much everybody in the world who is using the web and/or mobile phones; the

[7] http://www.whitehouse.gov//sites/default/files/microsites/ostp/Data2Action%20Announcements.pdf

> second group is smaller; and the third group is much smaller still. We can refer to these three groups as the new "data-classes" of our big data society.[8]

Finally, studying and responding to the legal ramifications associated with data acquisition, use, and security is a rapidly-growing area in domestic and international law.

Research Practices and Outcomes

The availability and volume of data, along with new ways of sharing information, are transforming how researchers conduct and communicate their work. This is a topic of immediate concern for scholars and scientists across disciplines. UNC currently has several high-profile projects involving expert multidisciplinary teams that address the disruptive potential for data in various fields of research. One aspect of working with data illustrated by these projects, and critical to student development, is the ability to form and function in cross-disciplinary, often multi-institutional teams with members of diverse skillsets and backgrounds. Another aspect worth noting is how information technology is changing the paradigm of the traditional publishing model in research.

An example of UNC's efforts to employ data to advance scientific discovery and economic competitiveness is the National Consortium for Data Science (NCDS), spearheaded by Stan Ahalt of UNC's Renaissance Computing Institute (RENCI). The NCDS was launched in April 2013 as a way to address the challenges and opportunities posed by massive data sets being created by digital medicine, environmental sensors, scientific instruments, social networks, and more. Its goals include: identifying key data science challenges; encouraging data science research that spans academia, industry and government; facilitating improved data science education; supporting technical, ethical and policy standards for data; and applying data science expertise to many societal problems and scientific disciplines, including genomics, environmental sciences, energy, sustainability, social and population studies, and materials science. NCDS founding members are RENCI, UNC-Chapel Hill, Cisco Systems, Drexel University, Duke University, GE, IBM, MCNC, the National Institute for Environmental Health Sciences, North Carolina State University, SAS, RTI International, Texas A&M University, UNC-Charlotte, UNC General Administration, and the U.S. Environmental Protection Agency.

Demonstrating the potential to harness data in fields beyond science, the Carolina Digital Humanities Initiative is transforming the model for traditional humanities researchers, with applications that range from curating online collections to mining information from large data

[8] L. Manovich, "Trending: The Promises and the Challenges of Big Social Data", Debates in the Digital Humanities, ed M.K. Gold. The University of Minnesota Press, Minneapolis, MN, 2011 (http://www.manovich.net/DOCS/Manovich_trending_paper.pdf)

sets. The initiative builds on the work of the Digital Innovation Lab led by Professor Robert Allen and seeks to improve all aspects of humanities research. The Digital Innovation Lab is currently working with the UNC Library and corporate partner Ancestry.com to explore the "big data" implications of the digitization of millions of pages of North Carolina newspapers currently being undertaken through an agreement between UNC and Ancestry.com subsidiary Newspapers.com.

Additional examples of the way that the volume and type of data available in various sectors have significant implications for research include:

- *Climate science*. Study of changes in the complex, interdependent systems that make up the earth's ecosystem must now inform real-time decision making in areas from city and regional planning, to water policy, to regulations regarding energy sources and consumption.

- *Healthcar*e. Fiscal concerns, combined with an increased focus on evidence-based medicine, are driving data analytics in the healthcare sector.[9] While the rate of adopting medical knowledge increases, the sheer volume of knowledge available is overwhelming. Today, a typical primary care doctor must stay abreast of approximately 10,000 diseases and syndromes, 3,000 medications, and 1,100 laboratory tests.[10] Even more fundamental is our existing concept of research. In the past, researchers only way of reliably understanding the world around them with any precision or accuracy was by using the scientific method. Now, in a world where one can directly observe the world via big data (though in an uncontrolled fashion), how do researchers establish a balance between the strengths of the scientific method and its inherent weaknesses? How do research methods adapt?

- *Urban development*. Basic urban functions such as transportation, education, infrastructure planning and maintenance, and policing can all be dramatically affected by a host of variables—including (but not limited to) changing demographics, economic stressors, and weather variations. Widespread technology adoption has now enabled collecting and real-time analysis of large new datasets to inform day-to-day decision making by planners and developers. NYU's Center for Urban Science + Progress uses urban informatics not only to collect vast datasets of information on New York City's

[9] "The big data revolution in US health care: Accelerating value and innovation," McKinsey, 2013
[10] B. S. Alper, J. A. Hand, S. G. Elliott, S. Kinkade, M. J. Hauan, D. K. Onion, and B. M. Sklar, "How much effort is needed to keep up with the literature relevant for primary care?" J. Med. Libr. Assoc.,vol. 92, no. 4, pp. 429–437, Oct. 2004

infrastructure and demographics, but to drive visionary planning based on an understanding of how people interact with their immediate environment.[11]

Impact on Industry and Organizations

Carolina students and faculty compete in a global marketplace. The arrival of the "big data" era is having an impact on organizations' human capital needs. For example, recruiters at the Kenan-Flagler Business School routinely ask about and screen for students' familiarity with data management tools and their analytical skills.

In a fall 2013 survey of UNC employers conducted by University Career Services, seven out of eight companies surveyed indicated that they expect managers and employees to be data literate, and that they seek more than just analytical tools. Employers are interested in new hires able to find, analyze and present insights gleaned from data. Half of the employers surveyed indicated their hiring process includes a means for screening applicants' "ability to collect, organize and present data." One respondent to the survey said: "Graduates should be strongly encouraged, if not required, to take more quantitative classes to better understand and interpret output from the big data work." Another respondent said: "In our undergrad recruiting, we find that some of our offerees don't have a ton of data analysis experience, which is fine (i.e. - not all of our offerees are business or econ majors, we hire a lot of liberal arts majors too). But they still must have the mind that can wrap around data in order to perform well."

Recognizing this potential gap in a student's skillset, a group of Deans Fellows at the Kenan Flagler Business School recently conducted an assessment of employee demand regarding data literacy (Appendix C). The group cited growing evidence of the likelihood that top-performing companies will require analytics experience in their new hires, along with increased demand from current MBA students for courses to increase both their understanding of data in context as well as their technical skills. The recommendation of the study is to make a concentration in Data Analytics and Decision Making available to MBA candidates.

"The Analytics Advantage," a survey conducted by Deloitte in 2013, reported that "96% of [corporate] respondents assert that analytics will become 'more important' or 'somewhat more important' during the next 3 years." Also, "nearly half of respondents (49%) assert that the

[11] http://cusp.nyu.edu/urban-informatics/

greatest benefit of using data analytics is that it is a key factor in better decision-making capabilities."[12]

Organizations collect growing amounts of data on markets they serve, along with data measuring the performance of processes, products, people and the entire enterprise. Data are used to improve performance through informing better decisions by leaders, managers and front line employees. The best decisions are evidence based and data driven. The challenge becomes what evidence and what data to use.

Technology has created a world where organizations are collecting more data than they can manage. Data is not only about scale – simply numeric data – but is "also characterized by the ability to render into data many aspects of the world that have never been quantified before; call it 'datafication.'" For example, location has been 'datafied' with GPS satellite systems. Words are treated as data when computers mine centuries' worth of books. Even friendships and 'likes' are 'datafied', via Facebook."[13] Big Data, a common term for the data sets now available and the challenges of managing them, is more than analytics. The Harvard Business Review[14] identifies three characteristics of Big Data that make it different from what managers are familiar with:

- Volume – the amount of data available doubles every 18 months.
- Velocity – much of the data being created has a shelf life; the value of the data quickly decays.
- Variety – big data takes the form of social media updates, sensor readings, customer feedback, performance evaluations, and much more.

A 2012 survey of senior Fortune 500 and federal agency business and technology leaders found that 85% of organizations surveyed had Big Data initiatives underway or planned.[15] This has significant implications for industry recruitment and training needs. The report stated, "What is less certain is how these same organizations plan to support these initiatives from a business and talent perspective." Forty percent of organizations reported it very difficult or impossible to find enough people with analytical skills in general and 66% found very difficult or impossible to find data scientists.[16]

[12] "The Analytics Advantage We're just getting stated," Deloitte, 2013
[13] Kenneth Neil Cukier and Viktor Mayer-Schoenberger,"The Rise of Big Data: How It's Changing the Way We Think About the World," Foreign Affairs, May/June 2013
[14] "Big Data: The Management Revolution," Harvard Business Review, October 2012
[15] Paul Barth and Randy Bean, "There's No Panacea for the Big Data Talent Gap," HBR Blog Network, November 29, 2012 (http://blogs.hbr.org/2012/11/the-big-data-talent-gap-no-pan/)
[16] "Big Data Executive Survey Themes & Trends," NewVantage Partners, 2012

A 2011 study at McKinsey Global Institute predicted that by 2018 there will be a 50% gap between demand and supply of deep analytical talent, amounting to a shortage of more than 150,000 data scientists.[17] Furthermore, the report stated, there will be an additional need for 1.5 million data literate managers and analysts with the skills to understand and make decisions based on the analysis of big data.

The Labor and Economic Analysis Division of the North Carolina Department of Commerce estimates that more than 18,000 data science jobs will be created in the state between 2010 and 2020, representing 4% of all newly-created jobs. Nearly all of these new positions will require a bachelor's degree or higher.

Given the interest from recruiters, the Faculty Working Group believes that nearly all firms and organizations hiring on campuses will expect candidates to be familiar with data techniques and comfortable with evidence based decisions. Among UNC employers surveyed, 100% agreed that employees at all levels of their organization "needed more data skills/experience."

According to the Harvard Business Review "… the advent of the big data era means that analyzing large, messy, unstructured data is going to increasingly form part of everyone's work. Managers and business analysts will often be called upon to conduct data-driven experiments, to interpret data, and to create innovative data-based products and services. To thrive in this world, many will require additional skills."

Corporate recruiters have begun to screen for these skills. For example Google's recruiters know that experimentation and testing are integral parts of their culture and business processes. So job applicants are asked questions such as "how many golf balls would fit in a school bus?" Or "how many sewer covers are there in Manhattan?" The point is not to find the right answer but to test the applicant's skills in experiment design, logic and quantitative analysis.

A 2012 survey by the business-technology firm Avanade estimated that knowledge workers spend 60% of their workday attempting to find and manage data.[18] The opportunity costs of hiring employees without data-literacy skills and comfort are large. Of UNC employers surveyed recently by UNC Career Services, only 1 of 8 agreed with this statement: "Employees hired out of college have data skills necessary for success in their organization."

A 2013 report from the American Management Association asserted that the top performing companies of the future will be those "with fully developed analytics skills – the ability to organize, analyze and communicate data that can be applied to their human capital and not just

[17] "Big Data: the next frontier for innovation, competition, and productivity," McKinsey Global Institute, June 2011
[18] "The challenge of big data: Turning data into knowledge and then putting it to work," Avanade, 2012.

to the other elements of their businesses." Employers are looking for employees with the skills to translate data into insights and to create value.

UNC RESPONSE

To establish a baseline for building data studies curriculum at UNC, the Faculty Working Group reviewed courses currently offered that address components of a targeted data-studies skillset. Reviews of the following curricula were completed (Appendix D):

- School of Public Health
- School of Information and Library Sciences
- College of Arts & Sciences Humanities Curriculum
- College of Arts & Sciences Natural Science Curriculum
- College of Arts & Sciences Social Science Curriculum
- Kenan Flagler Business School

This preliminary review showed that UNC currently offers numerous data-related courses at both the graduate and undergraduate level. The opportunity in developing a more focused and compelling data studies curriculum will be to connect existing expertise and courses, and identify and fill gaps to achieve learning outcomes consistent with a strong data studies skill set. New courses will likely need to be created that provide both a structural framework for integrating technical and non-technical aspects of data studies, as well as a contextual framework that takes advantage of UNC's strengths in the humanities and social sciences. As mentioned earlier, these courses will need to take into account the differing mathematical and computational skills across the schools and curricula. And, finally, the new focus on data studies and the curricular offerings will need to be well publicized and presented in a way that entices all populations of students.

For undergraduates, numerous courses that fulfill general education requirements in Quantitative Reasoning (QR), Quantitative Intensive (QI), Physical & Life Sciences, and Social & Behavioral Sciences contain components relevant to a data studies curriculum. The School of Information and Library Sciences already offers a minor in Information Systems, available to all undergraduates, which may have some components that intersect with data studies. A minor with a higher emphasis on technical skills, likely titled "Minor in Data Science," is being considered by the Department of Statistics and Operations Research, Math, Applied Physical Sciences and Computer Science.

At the graduate student level, the professional schools have the opportunity to create certificate programs (9 credit hours plus activity outside the coursework) in data studies or data sciences that specifically address the needs of their students. Certificates at the graduate level are often interdisciplinary and involve courses outside of one particular school. The graduate school offers support for course and certificate development. A graduate certificate in qualitative studies is already offered in the School of Education and is available to students in other graduate programs on campus. The main purpose of the certificate is to advance interdisciplinary qualitative inquiry through the knowledge, skills, and application of multiple qualitative tools and techniques. SILS offers a certificate in digital curation, the main goal of which is to teach students how to plan, manage and implement practices that ensure the long-term integrity and use of resources created in digital form. The certificate aims to prepare graduates for data-intensive jobs after graduation. New groupings of courses that intersect with data studies have been proposed in the Kenan Flagler Business School (Appendix E). This demonstrates both the recognition of student interest and need, as well as faculty interest and expertise for developing this topic area at UNC.

Identifying which of these courses exist and how they relate to one another is a first step in creating data studies curriculum. Of particular interest is creating cohesiveness and a framework to add focus and opportunities for cross-disciplinary learning.

Support from the University Libraries

The University Libraries have a well-established tradition of collecting, making accessible, and preserving data and datasets in support of both research and teaching. Library subject liaisons and the Data Librarian are available to assist faculty in developing courses and/or curricula that incorporate data studies, and can help connect faculty with suitable datasets for assignments or course modules. The Libraries have responded to the growing importance of data across all disciplines by making the support of the research lifecycle a central goal in their new strategic plan. A significant part of this aim will be realized through the creation of a Research Hub. The Research Hub will bring together experts in data, qualitative and quantitative analysis, GIS, data visualization, digital humanities and scholarship, and data management, to form a center of expertise available to Carolina's students, faculty, and staff.

Another service, available now, is the Carolina Digital Repository (CDR). The CDR offers long-term access and safekeeping for scholarly works, datasets, research materials, educational resources, university records, and audiovisual materials produced by the UNC Chapel Hill community. Open Educational Resources that are created to support curricula in data studies could be hosted and preserved by the CDR.

Curriculum Experts

The UNC Office of Undergraduate Curricula is available as a resource for faculty who wish to incorporate components of a data studies curriculum into their current courses, or for new course development. Funding for course revision ($2,500 - $3,000) or course development ($5,000 - $6,000) will help faculty to pursue incorporating data studies into the curriculum in a meaningful way.

Faculty have many paths to help equip UNC students with a data studies skill set. Incorporating teaching modules into existing courses, new course development, undergraduate minor, graduate certificates, majors and co-curricular programming (workshops, trainings, etc.) are all available mechanisms for widely adopting data studies into the student experience.

NEXT STEPS

In light of the Faculty Working Group's findings and the strong case for the need, these are the recommended steps for advancing data studies at UNC:

Share These Findings

To facilitate a campus-wide conversation, members of the Faculty Working Group will present the findings of this report to the Provost, deans, department chairs, and institute/center directors.

Ask Provost to Form a Steering Committee

Ask the Provost to form a small, faculty-led cross-campus Steering Committee to continue the work. It is recommended that its members be faculty who are passionate about the topic and committed to facilitating the adoption of data studies at UNC drawing in part from this Faculty Working Group. We also recommend that at least one student from each of the four subset student categories be included, as well as curricular experts from around the campus and university librarians. (See Appendix F for one student's input).

Areas of focus for this committee:

- Continue to circulate the findings of the Data Studies Working Group.
- Develop targeted and detailed curricular strategies for undergraduate and graduate audiences.
- Form a research and industry advisory group.

- Consider possible connections to curriculum in conversations regarding UNC's infrastructure and support for Big Data.
- Publicize and drive student awareness of the value of these curricular offerings.
- Conduct outreach to organizations and companies with a strong interest in data literacy and, in many cases, resources to help (i.e., IBM, Cisco, SAS, etc.)

Create Awareness

The Steering Committee may work with the Center for Faculty Excellence, which regularly convenes forums to raise awareness of trends and generate support for new curricular content—to host a campus-wide forum on Data Studies. Working Group members will encourage colleagues within their respective academic units to explore the relevance and implementation within their curriculum.

The Steering Committee could present at the quarterly meeting of the Office of Undergraduate Curricula, which brings together academic leads and has offered to add this topic to the agenda of its first fall meeting. Similarly, the Graduate School is available to work with academic leads on exploring ways to integrate data studies into existing curricula and develop new courses.

It is incumbent on the Steering Committee and those advocating for data studies curriculum to develop ways to actively encourage undergraduates and graduates to pursue the courses and opportunities of the curriculum. This could include creating accessible web resources for identifying relevant courses as well as providing resources to departments for creating research programs that involve undergraduates and graduate students in integrated projects involving data sciences and existing field-specific strengths. Some of this work has been started and is housed on the web site described below.

Build a Curricular Resource Center

Resources compiled by the Faculty Working Group are available at the website (https://datastudies.web.unc.edu), and include:

- Course inventories
- Course syllabi
- Benchmarking reports on undergraduate and graduate programs across the country that relate to data studies and/or data science
- Survey of UNC employers
- Articles and reports

BUDGET

Effecting meaningful and lasting change to UNC's curriculum will require both a modest financial commitment by the institution and a greater time commitment by faculty and staff. Although specific budget recommendations are outside the scope of this Faculty Working Group, we do encourage the recommended cross-campus Steering Committee to conduct a thorough assessment of the resources required to fully implement a robust data studies curriculum. Cost categories that might be considered include, but are not limited to: faculty release time for course development, teaching awards in data studies, data-specific student scholarships, marketing the courses, and personnel oversight.

CONCLUSIONS: THE ROAD AHEAD

At this stage the main question is not *whether* UNC should build up its curricula in data studies, but how to do so with excellence. The needs are evident, and we have tried to summarize them concisely in this report: data studies are essential to prepare our students for the new demands of the workplace, to contribute to their fields of study and society, and to provide them with the skills necessary for translational pursuits.

There may be some significant challenges on the road to building excellent curriculum components. For example, along with the goal of having data studies taught in all disciplines, the offerings in each discipline must be appropriate to that school or department (which can be achieved only through the buy-in and expertise of the faculty members concerned). Moreover, as has been emphasized in this report, work in data studies is inherently cross-disciplinary—so that factor must be addressed as well, in all aspects of curriculum-building.

Also, building and maintaining excellent curricula in data studies will entail aiming at moving targets. The age of "big data" is still young and evolving, with new applications or new possibilities emerging constantly. Ideally, in the course of building a curriculum, we at UNC-Chapel Hill will also be building the human and institutional capabilities to play an adaptive, forward-looking leadership role in this vitally important field.

The challenges are thus considerable but they are far from insurmountable. In the view of the Faculty Working Group, the key next step is the formation of a very strong Steering Committee. The members of this Committee will play multiple roles. They will be spokespersons and standard-bearers for data studies. They will be recruiters and early-stage coordinators of curriculum efforts among the faculty; they will be planners who must keep sharp eyes on needs and constraints.

Last but not least, the Steering Committee members must do their work in such a way as to turn it over to others, so that academic leaders and faculty across campus ultimately "own" the curriculum endeavor in every sense. Once that stage is reached, UNC will be well positioned as an institution ahead of the technological trends and a university that prepares its students for the world they will inherit.

https://datastudies.web.unc.edu

APPENDICES

APPENDIX A

UNC CHAPEL HILL FACULTY WORKING GROUP ON DATA STUDIES CURRICULUM MEMBERS

	Name	Title/Department	Unit	Email Address
1	Ahalt, Stan	Director, RENCI	RENCI	ahalt@renci.org
2	Allen, Robert	Distinguished Professor, American Studies	Arts & Sciences, American Studies	Robert_allen@unc.edu
3	Anderson, Daniel	Professor, Dept. Associate Chair	English & Comp literature	iamdan@unc.edu
4	Angle, Ray	Director, University Career Services	Student Affairs, Career Service	rayangle@email.unc.edu
5	Argon, Nilay Tank	Associate Professor, Dept. of Statistics	Arts & Sciences, Math & Statistics	nilay@unc.edu
6	Balasubramanian, Sridhar	Associate Dean, MBA Program	Business	drb@unc.edu
7	Barnwell, Stephanie	Programs Coordinator	Carolina digital humanities initiative	Stephanie.barnwell@unc.edu
8	Bolas, Michelle	Program Manager, Office of Innovation & Entrepreneurship	Chancellor's Office of I&E	Michelle.bolas@unc.edu
9	Brews, Peter	Professor & Associate Dean, KFBS	Business	Peter_brews@unc.edu
10	Caren, Neal	Assistant Professor, Sociology	Arts & Sciences, Sociology	Neal.caren@unc.edu
11	Carsey, Thomas	Distinguished	Arts & Sciences, Odum	carsey@email.unc.edu

		Professor, Political Science & Director, Odum Institute		
12	Cone, Judith	Special Assistant to the Chancellor for Innovation & Entrepreneurship	Chancellor's Office of I&E	Judith.cone@unc.edu
13	Cunningham, Robin	Lecturer, Statistics and Operations	Arts & Sciences, Math & Statistics	rjcunnin@email.unc.edu
14	Furey, Terry	Assistant Professor, Genetics & Biology	Medicine, Genetics	tsfurey@unc.edu
15	Hannig, Jan	Associate Professor, Statistics	Arts & Sciences, Math & Statistics	jan.hannig@unc.edu
16	Hunter, Carol	Deputy University Librarian	Library	cfhunter@email.unc.edu
17	Jones, Paul	Professor/Director of ibiblio.org, Journalism & Mass Communication	Information and Library Science	pj@unc.edu
18	Kelly, Kip	Director of Marketing, KFBS	Business	kellykm@email.unc.edu
19	Knowles, David	Director of Economic Development & Regional Engagement, RENCI	RENCI	davidknowles@unc.edu
20	Krishnamurthy, Ashok	Deputy Director, RENCI	RENCI	ashok@renci.org
21	Lach, Pam	Digital Innovation Lab Manager, American Studies	American Studies	plach@email.unc.edu
22	Lastra, Anselmo	Professor, Computer Science	Arts & Sciences, Computer Science	lastra@unc.edu

23	Leloudis, James	History Professor	Arts & Sciences, History	leloudis@unc.edu
24	Leousis, Peter	Deputy Director, Odum Institute	Arts & Sciences, Odum Institute	peter_leousis@unc.edu
25	Marciano, Richard	Professor, School of Information & Library Sciences	Information & Library Science	Richard_marciano@unc.edu
26	McKeen, Shannon	Exec. Director of Global Corporate Relations, KFBS	Business	Shannon_mckeen@kenan-flagler.unc.edu
27	Mizzy, Danianne	Head of Kenan Science Information Services, University Library	Library	mizzy@email.unc.edu
28	Mucha, Peter	Applied Physical Sciences, Chair; Distinguished Professor, Mathematics	Arts & Sciences, Applied Physical Sciences	mucha@unc.edu
29	Neta, Ram	Associate Professor, Philosophy	Arts & Sciences, Philosophy	neta@email.unc.edu
30	Nobel, Andrew	Professor, Statistics and Operations Research	Arts & Sciences, Math & Statistics	nobel@email.unc.edu
31	Pipiras, Vladas	Professor, Statistics and Operations Research	Arts & Sciences, Math & Statistics	pipiras@email.unc.edu
32	Rajasekar, Arcot (Raja)	Chief Scientist Professor	RENCI SILS	sekar@renci.org
33	Seidman, Rachel	Associate Director, Southern Oral History Program	Arts & Sciences, Center for the study of the South	seidman@email.unc.edu

34	Smith, Kennetha	Executive Assistant to Judith Cone and Michelle Bolas	Chancellor's Office of I&E	Kennetha@unc.edu
35	Thornburg, Ryan	Assistant Professor, Journalism/Mass Communication	Journalism & Mass Communication	Ryan.thornburg@unc.edu
36	Wildemuth, Barbara	Associate Dean and Professor, SILS	Information & Library Science	wildemuth@unc.edu

THE UNIVERSITY
of NORTH CAROLINA
at CHAPEL HILL

Data Science Education

Report, Summer 2013

Erin Morris
Kenan Science Library

A report on the current state of post-secondary education programs in data science and related disciplines in U.S. universities. The impetus for such programs is discussed, as are curriculum trends. Program requirements for a number of programs at both the graduate and undergraduate levels are summarized.

Contents

Introduction

Introduction 25
Undergraduate Course Topics 26
Selected Undergraduate Program Summaries 28
George Mason University (Data Analysis Minor) 28
College of Charleston (Data Science/Discovery Informatics Major) 28
College of Charleston (Discovery Informatics Minor) 28
Ohio State University (Computational Science & Engineering Minor) 28
Graduate Course Topics 29
Selected Graduate Program Summaries 31
George Mason University (Graduate Certificate in Data Analytics) 31
Northwestern University (M.S. in Analytics) 31
NYU (M.S. in Data Science) 31
References 32
Appendix I: Undergraduate Curriculum Overviews 33
Undergraduate Minors: Data Science/Informatics 33
Undergraduate Minors: Computational Science 34
Undergraduate Majors 37
Appendix II: Graduate Curriculum Overviews 39
Graduate Degrees 39
Graduate Certificates 42
Appendix III: All Programs 45
Undergraduate (Minor) 45
Undergraduate (Major) 45
Graduate Certificate 45
Graduate (M.S.) 46

Introduction

The 'Big Data' revolution has radically changed research practices in both industry and academia. Increasingly large and complex datasets have created a new paradigm for research: the current challenge is not collection of data, but rather extraction of knowledge from data (Djorgovski, 2005; Foster, 2006). A skilled workforce is needed to develop new computational methods to analyze, visualize, and preserve these massive amounts of data; a 2011 report by the McKinsey Global Institute projects a shortage of over 140,000 data scientists by 2018 (Brown et al, 2011). Given the steady growth in processing power and data set size, data storage and preservation are also concerns.

In response to this "data deluge," many universities have established degree or certificate programs in data science, analytics, or data management. These programs vary in level and scope, but most share a common core of skills, including statistical analysis, computation and programming, and communication. These curricula generally require support from multiple academic departments; not only is data science applicable to many disciplines, but data science itself spans several other established disciplines, primarily mathematics, computer science, and information science (Miller, 2013). Interdisciplinary elective offerings and capstone projects allow students to focus their studies on particular areas of interest. The programs surveyed for this report are focused on the practical applications of data science, rather than theory: many curricula stress the use of "real-world" data for coursework and projects, and several programs include optional industry practicums and internships.

Weber et al (2012) state that a key aspect of data management education is integration of knowledge from across the disciplines; data professionals should be familiar with the research processes that produce the data as well as the needs of the researchers that use it. Thus, a major challenge for designing a data science curriculum is maintaining currency with trends in research practice and advances in computational methods. A recent survey of academic science researchers identified three main job aspects for eScience professionals: managing and analyzing data, communicating with scientists and information technologists, and working with computers and software (Kim et al, 2011). From this focus group data, the authors compiled a "top ten list" of recommended course topics for data professionals, including data curation, data mining and database design, introductory web and software programming, cyberinfrastructure and information systems, and scientific research.

Undergraduate Course Topics

Data science and eScience are practiced at the interface of mathematics, computer science, and a scientific discipline. Most minor curricula seem designed to 'fill in the gaps' for students trained primarily in one of these areas, providing introductory training in statistics, programming, and research methods.

Data Mining and Databases

Data mining was included in all undergraduate data science curricula (majors and minors), and in a few of the computational science programs. A recent case study of an undergraduate data mining course showed that the knowledge gap between CS and non-CS majors created a challenge for syllabus designers. The authors suggested splitting the course into two sections: one a computer science elective with prerequisites in data structure and algorithms, and an introductory, interdisciplinary course for non-majors. The authors also stress the importance of statistics and probability as prerequisites for data mining courses. (King & Satyanarayana, 2013)

Information Ethics

Given the recent headlines about Big Data and privacy, information ethics is a worthwhile addition to a data science training curriculum. The Computational and Data Sciences major at George Mason requires an entire course in information ethics. Indiana University's Informatics minor offers an ethics elective. In programs without a dedicated course, ethics often appeared in course descriptions for introductory classes, as well as some advanced discipline-specific informatics courses.

Mathematics & Statistics

Minors were generally aimed towards students from the sciences and engineering, who often have some background in mathematics. Thus, basic calculus was rarely included in the core courses for minors, though some had extensive math prerequisites. Linear algebra and introductory probability & statistics were common core or prerequisite courses; optimization, modeling, and numerical analysis were common as both core courses and electives.

Programming

The surveyed undergraduate programs rarely required prior programming experience or programming prerequisites. Most minors included a basic programming class, either a general "Computer Science I" or an introductory "Programming and Computation for the Sciences" course.

Modeling and Visualization

Modeling was a common course topic in undergraduate curricula, either as an overview methods course or in the context of a discipline-specific elective (e.g. "molecular modeling"). Visualization was rarely overtly included in course titles, though it is presumably included in introductory data science courses.

Disciplinary Courses

Upper-level electives with a discipline-specific focus (e.g. 'Bioinformatics') were extremely common in data science majors/minors, and fairly frequent in computational science programs. These courses add a practical element to the curriculum and allow students to link their data science minor more closely with

their major, or with another discipline (e.g. a Chemistry major with expertise in genomic data analysis). Sometimes a major project is integrated into this course.

Capstone Project/Independent Study

These were less common in undergraduate programs than in graduate programs. A capstone project is a formal requirement of the computational science minor at Ohio State University and the data science major at the College of Charleston. Capstones, independent studies, and internships were available as electives in just under half of the surveyed minor curricula.

Selected Undergraduate Program Summaries

Minor programs of study ranged from 12 credits to 27 credits in length, with five to six courses being the most common. Majors required 60 -80 credits and usually consisted of about 8 data science courses, plus extensive requirements in statistics, computer science, and mathematics. The program at George Mason has prescribed specializations in Biology, Chemistry, or Physics; data science majors at the College of Charleston select a "cognate" from one of 14 focus areas in business and the sciences.

George Mason University (Data Analysis Minor)

This 15-credit undergraduate minor is intended for students in the STEM and health disciplines, to provide extra experience with statistics and data analysis methods. The minor includes 2 introductory classes in probability and statistics, and 3 additional courses. These electives may be either broad (such as "data mining," "database concepts," or "research methods") or subject-specific (e.g. "statistics in psychology," "GIS in engineering").

Suggested background: This minor is designed to "complement" science and engineering majors.

College of Charleston (Data Science/Discovery Informatics Major)

This 69-credit undergraduate major is administered by the Department of Computer Science. The core requirements consist of 2 introductory data science courses (including such topics as modeling, data mining, and database management) and a capstone project. The major also requires 45 credits of mathematics and computer science courses. Students select a "cognate" from one of 14 focus areas in business and the sciences; this involves about 5-6 courses (and sometimes additional labs) in a single subject area, and determines the focus of the capstone project.

College of Charleston (Discovery Informatics Minor)

This 20-credit undergraduate minor consists of 2 introductory data science courses, 2 programming courses, and 3 mathematics courses (pre-calculus and statistical methods). It is intended as a "broad overview" of data mining and informatics field, and is appropriate for any students wishing to augment their degree program with additional quantitative and analytical skills.

Suggested background: The core classes do not appear to have prerequisites so this minor should be appropriate for any major.

N.B.: The College of Charleston appears to be rebranding its data studies programs, with "Data Science" transitioning to "Discovery Informatics." However, this is not consistent across the entire website or course catalog.

Ohio State University (Computational Science & Engineering Minor)

This undergraduate minor is designed for science and engineering students seeking computational skills for application to their home disciplines. The program requires one course in each of six subject areas (simulation, optimization, numerical methods, programming, computation, and visualization) and a capstone project or internship.

Suggested background: Calculus prerequisites and the upper-level nature of many of the electives might make the minor inaccessible to students outside the hard sciences.

Graduate Course Topics

Databases, Data Mining, & Machine Learning

Every curriculum appears to have some component involving database management and querying, either in a focused class, or as a unit in the introductory data science course. Data mining and/or machine learning was a core topic in about 2/3 of the programs surveyed, and available as an elective for others. Columbia, Stanford, Northwestern, and NYU all offer courses with at least a titular focus on "Big Data," with topics including mining massive datasets, distributed datasets, and Hadoop (or other large-scale data software).

Statistical Methods, Modeling, & Visualization

Some programs are weighted more heavily towards computer science than numerical methods, but most include statistics in some form. The undergraduate and graduate programs at George Mason University, as well as the graduate analytics programs at Northwestern and NCSU, place a strong emphasis on statistics. Harvard's Computational Sciences program is also focused on mathematical methods. Statistical modeling and data visualization were included in the core curriculum of about half of the programs surveyed.

Programming, Software, & Computational Methods

Pure programming is not heavily emphasized in the core classes of these programs, though some (like Stanford, NYU, George Mason, and Illinois Tech) assume that students will enter with some programming experience. Cornell's Applied Statistics M.S. curriculum includes a Python programming course, and Columbia's "Algorithms for Data Science" course involves programming. One notable aspect of the M.S. in Computational Science at George Mason is the "High Performance Computing" course, which includes topics such as hardware and software design, and scalability for managing large datasets.

Communication & Project Management

Communication skills should be emphasized in a data science curriculum because of the niche that data professionals occupy in the research process. Data specialists often work as intermediaries between researchers and I.T. specialists, or researchers and the public, and training in data visualization and scientific research methods can assist them in these roles. The data programs at Northwestern University and N.C. State University place particular emphasis on teamwork and communication to a variety of audiences, and a course in "Communicating Science" is a core requirement for the M.A.S. program at Illinois Institute of Technology. IIT also requires a course in project management, one of the "top ten" topics suggested by Kim et al (2011).

Capstone Project/Practicum

A final project or practicum in a "real-world" setting provides highly relevant work experience; the most rigorous examples are in the programs at Northwestern, NYU, and NCSU. Northwestern's program includes a 12-month industry practicum followed by a one-semester capstone project (also with data from a "real" client).

Electives

Elective availability varies. The shorter certificate programs, like those at U. Washington, Columbia, George Mason, and Stanford, do not include any electives. Longer programs offer a wide array of electives through cross-listing with other departments. George Mason's Computational Sciences M.S. program includes many high-level electives in mathematics and the physical sciences. The M.S. program at NYU includes 18 elective hours and 11 pre-approved specialization areas, including Bioinformatics, Artificial Intelligence, Sociology, and Business.

Selected Graduate Program Summaries

George Mason University (Graduate Certificate in Data Analytics)

The Data Analytics Graduate Certificate is a 12-credit, part-time program designed for professionals working in fields affected by Big Data (healthcare, business, bioinformatics, etc.). Courses provide an overview of Big Data, including ethics/policy issues, analytics, and data use in various spheres. Other major topics include database structure, operations/decision research, and statistical methods.

Suggested undergraduate background: computer science, STEM, I.T., or data/analytics.

Northwestern University (M.S. in Analytics)

This full-time, 15-month program is administered through the School of Engineering, with additional teaching support from faculty in management, marketing, communications, and computer science, as well as industry practitioners. The curriculum covers "all three types of data analysis" (forecasting, data mining, and simulation). Hands-on industry experience is emphasized through an 8-month industry practicum, summer internship, and capstone project; students are also trained in industry software and tools (SPSS, SAS, etc.). Data management and Big Data analytics are taught in the 1st and 3rd semesters, respectively.

Suggested background: science, engineering, or business.

NYU (M.S. in Data Science)

The NYU Center for Data Science is a collaborative effort from many departments across the university, and the 36-credit M.S. degree program will launch during Fall semester 2013. The required courses provide an introduction to statistical analysis, machine learning, and Big Data. Half of the required credits are electives, which may be chosen from 11 pre-approved subject areas, from STEM disciplines to the humanities and health sciences. This flexibility in the curriculum, as well as the capstone project, should allow a high level of degree customization.

Suggested background: Some math (linear algebra and calculus) and programming experience.

References

Atkins, D. et al. (2003). Revolutionizing science and engineering through cyberinfrastructure: report of the National Science Foundation blue-ribbon advisory panel on cyberinfrastructure.

Brown, B.; Bughin, J.; Byers, A. H.; Chui, M.; Dobbs, R.; Manyika, J.; Roxburgh, C. (2011). Big data: The next frontier for innovation, competition, and productivity. McKinsey Global Institute. Retrieved from http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation.

Djorgovski, S. G. (2005). Virtual Astronomy, Information Technology, and the New Scientific Methodology**.** http://arxiv.org/abs/astro-ph/0504651

Foster, I. (2006). 2020 Computing: A two-way street to science's future. *Nature 440,* 419 (23 March 2006). doi: 10.1038/440419a

Kim, Y.; Addom, B.K.; Stanton, J.M. (2011). Education for eScience professionals: integrating data curation and cyberinfrastructure. *The International Journal of Digital Curation 6 (1),* 125-138. Doi: 10.2218/ijdc.v6i1.177

King, B. R., & Satyanarayana, A. (2013). Teaching Data Mining in the Era of Big Data. *Proceedings from the 120th ASEE Annual Conference and Exposition.* Atlanta, GA.

Miller, C. C. (2013, April 14). The numbers of our lives. *The New York Times*. Available from LexisNexis Academic, http://www.nytimes.com/2013/04/14/education/edlife/universities-offer-courses-in-a-hot-new-field-data-science.html

Appendix I: Undergraduate Curriculum Overviews

Undergraduate Minors: Data Science/Informatics

College of Charleston

Program	Discovery Informatics
Level	Undergraduate (minor)
Length	6 courses; 20 credits
Core Classes	Introduction to DS (modeling, stat interference, data mining) Dataset organization and Management (Relational DB structure, storage) Computer Programming I (+1-hour lab) Pre-Calculus Statistical Methods I/II
Electives, etc.	N/A
Notes	The College of Charleston appears to be rebranding its data studies programs, with "Data Science" transitioning to "Discovery Informatics." However, this is not consistent across the entire website or course catalog.

George Mason University

Program	Computational and Data Sciences Minor
Level	Undergraduate (minor)
Length	5 courses; 15 credits
Core Classes	Intro to Computational & Data Sciences OR Computing for Scientists
Electives, etc.	+ Any upper-level science course + 3 courses in Computational Science & Informatics **OR** Data Science departments
Notes	"…for students majoring in mathematics, science, or engineering who wish to augment their major degree program with additional courses in scientific computing."
Program	Data Analysis Minor
Level	Undergraduate (Minor)
Length	5 courses; 15 credits
Core Classes	6 credits of probability and stats
Electives, etc.	+ 9 credits electives, topics include: Data mining Databases Statistical analysis Modeling Discipline-specific informatics
Notes	From the Statistics department, to complement Stats, Econ, Engineering, etc. "The minor provides students with a background in data analysis and statistical methodology. It is intended to complement undergraduate degree programs in the Volgenau [Engineering] School and the College of Science…"

Illinois Institute of Technology

Program	Specialization in Data Science
Level	Undergraduate (within Computational Science/Information Systems majors)
Length	4 courses
Core Classes	Intro to Distributed Computing Data Mining **OR** Machine Learning Math: Stochastic Processes **OR** Design & Analysis of Experiments Business: Strategies for reaching new markets
Specializations	N/A
Electives, etc.	N/A
Notes	"Computational and statistical skills to build systems that can generate meaningful insights from massive amounts of data."

Indiana University

Program	Informatics Minor
Level	Undergraduate (minor)
Length	5 classes; 17 credits
Core Classes	Introduction to Informatics Programming (choice of 3)
Electives, etc.	+ 3 upper-level Informatics electives: Disciplinary informatics electives Capstone Project Data Mining Ethics Internship
Notes	Within the School of Computer Science and Informatics

University of Mary Washington

Program	Undergraduate Minor in Data Science
Level	Undergraduate (minor)
Length	7 classes; 23 credits
Core Classes	Intro to Statistics Linear Algebra Computer Science I Data Mining Modeling and simulation
Electives, etc.	Data Structures **OR** Analytics Application Development Foundations/Applications of Data Analytics **OR** Parallel Computing
Notes	Collaboration between Math, Computer Science, and Business departments.

Undergraduate Minors: Computational Science

Vanderbilt

Program	Scientific Computing
Level	Undergraduate (minor)
Length	5 classes; 15 credits
Core Classes	Introductory Programming

	Program Design Intro to Scientific Computing (disciplinary examples, data mining, visualization)
Electives, etc.	+ 2 electives (course list): High performance computing **OR** Independent Study **OR** discipline-specific upper-level courses (varied, but modeling is common)
Notes	Disciplinary courses in Astronomy, Biology, Biomedical engineering, Chemistry, Math, Mechanical Engineering, Physics, Psychology

Miami University (OH)

Program	Computational Science & Engineering Minor
Level	Undergraduate (minor)
Length	18-27 hours, 8 courses
Core Classes	Calculus (9) Simulation & Modeling for science/engineering Numerical Analysis Intro Programming Optimization
Electives, etc.	+ 1 discipline-related elective + High-performance computing **OR** Differential Equations
Notes	"This minor provides students majoring in science and engineering with specific skills to complete computationally based projects. These skills, which are highly sought by employers, include mathematical methods, computer programming, and computational modeling and simulation"

Ohio State University

Program	Computational Science minor
Level	Undergraduate (minor)
Length	6 courses + capstone/internship, ~19-22 credits
Core Classes	Simulation & Modeling Programming & Algorithms Numerical Methods Optimization
Electives, etc.	+ Discipline specific elective and capstone + Elective (visualization OR differential equations)
Notes	Math prerequisites (Calculus). "…students who already have expertise in science and engineering skills complete computationally based projects"

University of Wyoming

Program	(Proposed) Undergraduate Minor in Computational Science
Level	Undergraduate (minor)
Length	5 classes; 15 credits
Core Classes	At least 2 of: Numerical Analysis High-Performance Computing Scientific Computing Statistical Computing and Modeling
Electives, etc.	Electives:

	Computational Biology Algorithms and Data Structures Mathematical and Computational Methods in Physics Molecular Modeling C with Numerical Methods for Engineers Mathematical Modeling Introduction to Finite Element Methods Principles of Database Systems
Notes	9 credits must be upper-level; 6 credits must be outside major.

Chapman University

Program	Undergraduate Minor in Computational Science
Level	Undergraduate (minor)
Length	9 classes, 27 credits (**4 classes/12 credits** distinct from math requirements)
Core Classes	1 of: Foundations of Scientific Computation Computer Science II Computational Mathematics Tools
Electives, etc.	3 Electives: Intro to Bioinformatics Data Structures and Algorithms Differential Equations Database Management Complex Analysis Numerical Analysis
Notes	15 credits of math (linear algebra and calculus) and computer science requirements, followed by 4 upper-level electives.

American University

Program	Computational Mathematics minor
Level	Undergraduate (minor)
Length	7 classes, 24 credits (12 credits must be unique to the minor)
Core Classes	Introduction to Computer Science (7) Tools of Scientific Computing (3) Introduction to Simulation & Modeling (3) Calculus (8) Linear Algebra (3)
Electives, etc.	N/A
Notes	"The minor in computational mathematics is designed to augment the education of any student of the sciences by giving that student fluency in computation."

Stanford University

Program	Mathematical and Computational Science minor
Level	Undergraduate (minor)
Length	Unclear from site; possibly 8 courses @ 3-4 credits each.
Core Classes	1-2 courses in **each** of: Computer Science Mathematics, Management Science/Engineering, Statistics

Electives, etc.	+ 3 electives from 2 constituent departments; topics include: Computing & Algorithms Data mining/analysis Modeling/optimization Theoretical math Applied Statistics Statistical Interference
Notes	

Undergraduate Majors

College of Charleston

Program	Data Science
Level	Undergraduate (major)
Length	69 credits: 3-class DS core, 26 credits Math and 19 credits CS prerequisites
Core Classes	Introduction to DS (modeling, stat interference, data mining) Dataset organization and Management (Relational DB structure, storage) Capstone (project in "cognate" [focus] area)
Electives, etc.	+ 45 credits Math/statistics and computer science requirements; e.g. statistics, programming, data mining, data structures + 15-23 credits from one of: Accounting, Biomechanics, Customer Relationship Management, e-Commerce, Economics, Exercise Physiology, Finance, Geoinformatics, Molecular Biology, Organismal Biology, Physics and Astronomy, Psychology, Sociology and Supply Chain Management
Notes	The College of Charleston appears to be rebranding its data studies programs, with "Data Science" transitioning to "Discovery Informatics." However, this is not consistent across the entire website or course catalog.

George Mason University

Program	B.S. Computational and Data Sciences
Level	Undergraduate (Major)
Length	~80 credits. 6 classes in CDS core.
Core Classes	Scientific Data and Databases Data Mining Modeling Scientific programming, computation, and data visualization
Specializations	Biology, Chemistry, or Physics
Electives, etc.	+ Information Ethics + Physics & Math reqs + 1-3 CDS electives Ex. Visualization, Sci. Programming, Independent Study
Notes	"Graduates of the BS program in CDS will possess the mathematical, scientific, and computational skills necessary to participate effectively as members of the interdisciplinary scientific simulation and analysis groups"

University of Michigan

Program	Informatics (Data Mining & Information Analysis track)
Level	Undergraduate (major)
Length	~60 credits; ~16 classes
Core Classes	Discrete Math Programming/Data Structures OR Information System Design/Programming Intro to Quantitative Research Methods
Electives, etc.	+ Track Ex. Data Mining & Information Analysis Track: Core: 15-16 credits; 4 courses Linear Algebra Intro Statistical Computing Data Mining and Statistical Learning PLUS: one course in Stats/Optimization/Programming + 3-4 electives; 12-13 credits Topics: Computing, Bioinformatics, Adv. Stats/Probability, Data Mining, HCI
Notes	Interdisciplinary major housed in the College of Literature, Science, and the Arts in cooperation with the College of Engineering and the School of Information. **Prerequisites**: 5 courses incl. Intro Programming, Calculus, Stats, Information Systems.

Appendix II: Graduate Curriculum Overviews

Graduate Degrees

Illinois Institute of Technology

Program	Professional Master of Data Science
Level	Graduate (M.S.)
Length	33 credits
Core Classes	18 credits Applied Stats Machine Learning Data processing Data prep & analysis Communicating Science Project management
Specializations	No formal tracks
Electives, etc.	+ 6 credit capstone + 9 credits of electives from Math or Computer Science: http://iit.edu/csl/programs/professional_masters/ds_academics.shtml
Pre-reqs	B.S.; coursework/experience in linear algebra/statistics, programming & databases.
Notes	Launching Fall 2013. "IIT's MAS DS program focuses on acquiring deep theoretical knowledge and understanding of the assumptions made by machine learning methods; object-oriented **programming** skills; **understanding of scientific problems and processes**; and skill in facilitating **effective communication with scientific collaborators**"

George Mason University

Program	Computational Sciences M.S.
Level	Graduate (M.S.)
Length	30 credits
Core Classes	3 of: Numerical Methods (computational techniques for science problems) Foundations of Computational Science (mathematical modeling) High-Performance Computing (hardware and software design & scalability) Scientific & Statistical Visualization Scientific databases (query languages and DB management for scientific data)
Specializations	
Electives, etc.	+ 15 credits from Computational Science/Informatics course list: + 6 credits science electives **OR** MS project.
Pre-reqs	Math and programming coursework/experience
Notes	From the School of Physics, Astronomy, and Computational Sciences/Informatics. "solid foundation in information technology skills with computational courses in a variety of scientific areas" Optional masters project

NCSU

Program	M.S. Analytics
Level	Graduate (M.S.)
Length	3 semesters, 10 months
Core Classes	Analytics Tools & Techniques (DB, querying, stats programming, linear algebra, data visualization) Analytics Foundations (Linear regression, tables analysis, stats assessments) Analytics Methods & Applications (Data mining, programming/simulation, modeling) Analytics Methods & Applications II (modeling, Big Data) Analytics Practicums
Electives, etc.	Integrated curriculum: http://analytics.ncsu.edu/?page_id=123
Notes	Since 2007. "MSA students learn by tackling genuine analytics problems with data provided by industry and government sponsors, and using leading tools such as SAS. The MSA has a practical orientation. It does not dwell on theory and is not intended to be a prelude to doctoral study." Focus on teamwork & communication skills.

NYU

Program	M.S. in Data Science
Level	Graduate (M.S.)
Length	36 credits; 3 or 4 semesters
Core Classes	Intro to DS Statistical and Mathematical Methods (probability, stats, optimization) Big Data (distributed databases, automated knowledge extraction) Machine Learning/Comp. Stats (pattern recognition, algorithmic aspects of MachLearn) Interference & Representation (graphical models, advanced statistical machine learning) Capstone Project ("real-world" data problems from industry/gov't/academia)
Electives, etc.	6 electives; Focus Areas/Departments: Artificial Intelligence Large-scale computing Statistical/Mathematical Methods Mathematical foundations Applied Statistics Economics Biology/Bioinformatics Business Mathematical Finance Neuroscience/Psychology Sociology
Prereqs	Calculus, statistics, and linear algebra, some programming
Notes	

Northwestern

Program	M.S. in Analytics
Level	Graduate (M.S.)
Length	16 courses + internship; 15 month full-time
Core Classes	
Electives, etc.	Integrated Curriculum: http://www.analytics.northwestern.edu/curriculum-and-

	career-prospects/Course_description.html
Notes	"…combines mathematical and statistical study with instruction in advanced computational and data analysis. Students learn to identify patterns and trends; interpret and gain insight from vast quantities of structured and unstructured data; and communicate their findings in practical, useful terms."

Program	M.S. in Predictive Analytics
Level	Graduate (M.S.)
Length	11 courses, online part-time
Core Classes	DB design Data warehouse & data mining Statistical Analysis Analytics & Data Collection Predictive Modeling I & II Project Management
Electives, etc.	+ Leadership + Capstone **OR** Thesis + 2 Electives (Advanced modeling **OR** Advanced analytics course)
Notes	"…combines mathematical and statistical study with instruction in advanced computational and data analysis. Students learn to identify patterns and trends; interpret and **gain insight from vast quantities of structured and unstructured data**; and **communicate their findings** in practical, useful terms."

Harvard

Program	M.S. in Computational Science and Engineering
Level	Graduate (M.S.)
Length	8 courses, 2 semesters
Core Classes	Computing Foundations Scientific Computing: Numerical Methods Systems development for Computational Sci Stochastic Optimization Methods
Electives, etc.	+ 4 Computer Science **OR** Applied Math electives :
Notes	"rigorous training in the mathematical and computing foundations of CSE. Complementing the foundational coursework will be independent research projects and elective courses focusing on the application of computation to one or more domains" computation-heavy: http://iacs.seas.harvard.edu/academic-offerings/courses

Cornell

Program	MPS in Applied Statistics (Data-Centered Track II)
Level	Graduate (Master of Professional Studies)
Length	30 credits; 2 semesters with sufficient academic background; 3 including prerequisites

Core Classes	SAS Programming Statistical methods Database management Python programming Visualization, modeling, and analysis Data analysis project (in-depth, full year)
Electives, etc.	Applied statistics courses in various departments http://www.stat.cornell.edu/mps/academic.html
Notes	Also: Introduction to Data Science (iSchool course)

Graduate Certificates

Syracuse

Program	CAS in Data Science
Level	Graduate (Certificate)
Length	15 credits; 5 courses
Core Classes	Database management Data Science Data curation
Specializations	Analytics Data Storage/Management Data Visualization General Systems Management
Electives, etc.	http://coursecatalog.syr.edu/2012/programs/data_science
Notes	Science Data Information Literacy Project: http://sdl.syr.edu/

University of Washington

Program	Certificate in Data Science
Level	Graduate (Certificate)
Length	9 credits, 3 courses, ~9 months
Core Classes	Introduction to Data Science (DB types, SQL, file types) Methods for Data Analysis (stats, experimental design, machine learning) Deriving knowledge from data at scale (advanced models & practical applications)
Specializations	N/A
Electives, etc.	http://www.pce.uw.edu/certificates/data-science.html
Notes	Offered both online and in-person. **MOOC**: Intro to Data Science: http://escience.washington.edu/blog/data-science-mooc-engages-students-solving-problems-real-organizations

George Mason University

Program	Data Analytics Graduate Certificate
Level	Graduate (Certificate)
Length	12 credits; 4 courses

Core Classes	Analytics: Big Data to Information (collection and processing, ethical/policy issues) Principles of Data Management and Mining (overview, DBs, mining, Hadoop) Analytics and Decision Analysis (prescriptive and predictive. OR, optimization, modeling, decision analysis) Applied Statistics and Visualization for Analytics (multivariate regression, data access, statistical graphics – using R software).
Specializations	N/A
Electives, etc.	N/A
Notes	

Columbia

Program	Certificate of Professional Achievement in Data Sciences
Level	Graduate (Certificate)
Length	12 credits; 4 courses, 2 semesters
Core Classes	Algorithms for Data Science (searching/sorting, programming, algorithms, Big Data) Prob & Statistics (calculus based S&P) Machine Learning for DS (intro. Methods & problems related to Big Data) Exploratory Data Analysis and Visualization (exploration and visual presentation of datasets)
Electives, etc.	Core (above); no electives
Notes	Fall 2013. (Master's program to follow 2014) On campus, with expectation of future online offerings. Part –time/evening schedule. **Additional classes:** Intro to Data Science: http://www.columbia.edu/cu/bulletin/uwb/subj/STAT/W4242-20133-001/ Applied Data Science: http://www.columbia.edu/cu/bulletin/uwb/subj/STAT/W4249-20131-001/

Program	Core Statistics Certificate
Level	Graduate (Certificate)
Length	12 credits; 4 x 8-week online course.
Core Classes	Algorithms for Data Science (searching/sorting, programming, algorithms, Big Data) Prob & Statistics (calculus based S&P) Machine Learning for DS (intro. Methods & problems related to Big Data) Exploratory Data Analysis and Visualization (exploration and visual presentation of datasets)
Electives, etc.	Core (above); no electives
Notes	Fall 2013. (Master's program to follow 2014) On campus, with expectation of future online offerings. Part –time/evening schedule.

Stanford

Program	Mining Massive Data Sets certificate
Level	Graduate (Certificate)
Length	4 courses; 1-3 years (part time)
Core Classes	Social & Information Network Analysis

	Machine Learning Mining Massive Data Sets Information Retrieval & Web Search
Electives, etc.	http://scpd.stanford.edu/public/category/courseCategoryCertificateProfile.do?method=load&certificateId=10555807#course-selections
Prereqs	"Applicants should have knowledge of basic computer science principles and skills, at a level sufficient to write a reasonably non-trivial computer program. Familiarity with algorithms, data structures, basic probability theory, and linear algebra. Useful background also includes work in computer systems, artificial intelligence, statistics, and database systems. "
Notes	Mixed – online and in person
Program	Data Mining & Applications Certificate
Level	Graduate (Certificate)
Length	9 credits (3 courses, 1-2 years)
Core Classes	Data Mining & Analysis AND 2 of: Paradigms for Computing with Data Modern Applied Statistics: Learning Modern Applied Statistics: Data Mining
Electives, etc.	N/A
Prereqs	"A background in probability and matrix algebra (minimum of one course at the undergraduate level or above)"
Notes	

University of North Carolina – Charlotte

Program	(Proposed) Certificate in Data Science and Business Analytics
Level	Graduate (Certificate)
Length	5 classes; 15 credits
Core Classes	Big Data Analytics for Competitive Advantage (3) Data Base Systems (3) Business Intelligence and Analytics (3)
Electives, etc.	+ 2 electives: Visual Analytics OR Network Science Decision Modeling OR Business Analytics class
Notes	Programming and math experience required.

Appendix III: All Programs

Undergraduate (Minor)

American University:

Chapman University

George Mason:

Computational and Data Sciences; 5 classes; 15 credits - http://catalog.gmu.edu/preview_program.php?catoid=22&poid=20403&returnto=4546

Data Analysis; 5 classes; 15 credits - http://catalog.gmu.edu/preview_program.php?catoid=22&poid=20454&returnto=4546

Illinois Institute of Technology: Data Science specialization (for computer/information science); 4 courses, 12 credits - http://www.iit.edu/csl/cs/programs/undergrad/data_science_undergrad.shtml

University of Mary Washington

Miami University (Ohio):

Ohio State University

Stanford University:

Vanderbilt University:

University of Wyoming

Undergraduate (Major)

College of Charleston: B.S. in Data Science, 69 credits - http://compsci.cofc.edu/academics/undergraduate-programs/data-science.php

George Mason University: B.S. in Data Science; 80 credits - http://catalog.gmu.edu/preview_program.php?catoid=19&poid=17704

University of Michigan

Graduate Certificate

Columbia:

Certificate of Professional Achievement in Data Science; 12 credits: http://idse.columbia.edu/certification

Core Statistics Certificate; 12 credits - http://quants.columbia.edu/statistics/certificate-program-overview

George Mason: Data Analytics Graduate Certificate; 12 credits - http://catalog.gmu.edu/preview_program.php?catoid=22&poid=21887&returnto=4546

Stanford:

Mining Massive Datasets Certificate; 12 credits - http://scpd.stanford.edu/public/category/courseCategoryCertificateProfile.do?method=load&certificateId=10555807#course-selections

Data Mining & Applications Certificate; 9 credits - http://scpd.stanford.edu/public/category/courseCategoryCertificateProfile.do?method=load&certificateId=1209602#course-selections

Syracuse: CAS in Data Science; 15 credits - http://ischool.syr.edu/future/cas/datascience.aspx

University of North Carolina – Charlotte:

University of Washington: CAS in Data Science; 9 credits - http://www.pce.uw.edu/certificates/data-science.html

Graduate (M.S.)

Cornell: M.P.S. in Applied Statistics (Data Track); 30 credits - http://www.stat.cornell.edu/mps/options.html

George Mason: Computational Sciences M.S.; 30 credits - http://catalog.gmu.edu/preview_program.php?catoid=19&poid=17706&returnto=4105

Harvard: M.S. in Computational Sciences & Engineering; 24 credits - http://iacs.seas.harvard.edu/master-of-science-in-cse

Illinois Institute of Technology:

M.S. Data Science; 9 classes + 6-hour capstone; 33 credits - http://iit.edu/csl/programs/professional_masters/data_science.shtml

Specialization (within M.S. Computer Science); 4 courses; 12 credits - http://www.iit.edu/csl/cs/programs/grad/mcs_da.shtml

NYU: M.S. in Data Science; 36 credits - http://cds.nyu.edu/academics/ms-in-data-science/curriculum/

NCSU: M.S. in Analytics; 3 semesters - http://analytics.ncsu.edu/?page_id=1799

Northwestern:

M.S. in Analytics; 16 courses + internship; 15 months - http://www.analytics.northwestern.edu/overview/index.html

M.S. in Predictive Analytics; 11 online courses - http://www.scs.northwestern.edu/program-areas/graduate/predictive-analytics/

UNC CHAPEL HILL COURSES RELATED TO DATA STUDIES
FACULTY WORKING GROUP ON DATA STUDIES CURRICULUM

Data Studies current course inventory: Humanities and Arts
May 22, 2013
Ryan Thornburg
School of Journalism & Mass Communication

Methodology:

- Searched humanities and arts department websites for word "data".
- Browsed course descriptions in Undergraduate Bulleting and department websites for word "data".
- Searched any syllabuses linked from department websites for word "data".
- Searched and browsed the Undergraduate Bulletin and department sites for "compu*" and "analy*" and "quant*".

African and Afro-American Studies
395 Field Research Methods in African Studies (3). Recommended preparation, AFRI 101. This course will prepare students to conduct field research in Africa by looking at how to write a proposal, how to get research permission, and how to collect qualitative data.

American Studies
None.

Art
None.

Asian Studies
None.

Classics
None.

Communication Studies
610 Reading Quantitative Research in Communication Studies (3). Permission of the instructor for nonmajors. Review of the basics of quantitative research (e.g., scientific method, modes of data collection, instrument development, data analysis techniques) with the goal of gaining skill in reading published articles in communication studies journals.

Dramatic Art
None.

English and Comparative Literature
None.

Germanic and Slavic Languages and Literature
None.

History
None

Music
51 First-Year Seminar: The Interplay of Music and Physics (PHYS 51) (3). Acoustics and music from a practical standpoint. (PX)

337 "What's in a key? Modal types and tonal practices in music of the Classic era"

Philosophy
155 Introduction to Mathematical Logic (3). Introduces the theory of deductive reasoning, using a symbolic language to represent and evaluate patterns of reasoning. Covers sentential logic and first-order predicate logic. (QR)

356 Topics in Mathematical Logic (3). Prerequisite, PHIL 155. Topics may include the predicate calculus with identity and the metalogic of formal systems, modal logic, decision theory, alternative logics, probability, and induction. (QR)

357 Induction, Probability, and Confirmation (3). Current accounts of evidence and observation, the confirmation of scientific theories, the logic of inductive reasoning, and the metaphysics and epistemology of chance. (QI)

455 Symbolic Logic (LING 455) (3). Introduction for graduates and advanced undergraduates not taking the PHIL 155–356 sequence. (QR)

Religious Studies
None.

Romance Languages and Literature
679 Spanish Pragmatics (3). Prerequisite, SPAN 377. This course is an introduction to the study of meaning and language use, with a focus on Spanish. Includes discussion of the classical texts in the field as well as analysis of a variety of data (corpora, fieldwork, and experimental materials).

Women's and Gender Studies
None.

Data Studies Current Course Inventory: Natural & Mathematical Sciences
May 22, 2013
Peter Mucha
Department of Mathematics

Methodology:

> Importantly, this inventory is very much influenced by my own (sometimes quite limited) understanding of the role of data in each field. For most Departments, I have attempted to err on the side of including courses that we might later decide do not have sufficient data content; but even then, I very likely missed some. In contrast, I tried to focus on selected course offerings from CS and STOR, since the majority of their courses might make it onto a list that errs on the side of inclusion. – PJM

APPL
APPL 430 Digital Signal Processing I (3). Prerequisite, COMP 110 or 116. This is an introduction to methods of automatic computation of specific relevance to biomedical problems. Sampling theory, analog-to-digital conversion, and digital filtering will be explored in depth.

Biology
57 First-Year Seminar: Detecting the Future: Human Diseases and Genetic Tests (3). Restricted to first-year students. A first-year seminar focusing on the future of human diseases and genetic tests.

423L Laboratory Experiments in Genetics (4). Prerequisite, BIOL 205. Experiments using a range of organisms—from bacteria to Drosophila, higher plants, and man—to sample organismal and molecular genetics. One lecture hour, four laboratory hours.

425 Human Genetics (GNET 425) (3). Prerequisite, BIOL 202. Permission of the instructor for students lacking the prerequisite. Pedigree analysis, inheritance of complex traits, DNA damage and repair, human genome organization, DNA fingerprinting, the genes of hereditary diseases, chromosomal aberrations, cancer and oncogenes, immunogenetics and tissue transplants. Three lecture hours a week.

452 Mathematical and Computational Models in Biology (MATH 452) (4). Prerequisites, BIOL 201 and 202, MATH 231, and either MATH 232 or STOR 155. Permission of the instructor for students lacking the prerequisites. This course will introduce analytical, computational, and statistical techniques, such as discrete models, numerical integration of ordinary differential equations, and likelihood functions, to explore topics from various fields of biology. Laboratory is included.

454 Evolutionary Genetics (3). Prerequisites, BIOL 201 and 202. Permission of the instructor for students lacking the prerequisites. The roles of mutation, migration, genetic drift, and natural selection in the evolution of the genotype and phenotype. Basic principles are applied to special interest topics. Three lecture hours a week.

463 Field Ecology (4). Prerequisite, BIOL 201. Application of ecological theory to terrestrial and/or freshwater systems. Lectures emphasize quantitative properties of interacting population and communities within these systems. Required laboratory teaches methodology applicable for analysis of these systems. Projects emphasize experimental testing of ecological theory in the field. Two lecture and six field hours a week.

525 Computational Analyses and Resources in Genomics (4). Prerequisites, BIOL 202, COMP 116, and STOR 155. Permission of the instructor for graduate students. Computational techniques for the analysis of large-scale genomics data. Databases and online genomic resources. Programming for standard file processing and development of analysis pipelines. Course includes a computational laboratory.

526 Computational Genetics (4). Prerequisite, BIOL 202. Permission of the instructor for students lacking the prerequisite. Honors course. Prior computer science and statistics coursework recommended. A study of the fundamental concepts underlying DNA/protein alignment, gene finding, expression analysis, genetic mapping, phylogenetics, and population genetics. Includes a one-hour laboratory.

527 Seminar in Quantitative Biology (3). Prerequisites, COMP 114 and MATH 232 or 283. Permission of the instructor for students lacking the prerequisites. Special topics in quantitative biology for advanced students. The course counts as a quantitative biology course for the major.

527L Laboratory Seminar in Quantitative Biology (1). Laboratory in quantitative biology for advanced students. The laboratory will involve mathematical analysis and modeling of biological systems and processes.

528 Systems Biology of Genetic Regulation (4). Prerequisites, BIOL 202, COMP 116, and MATH 232 or 283. The course will focus on mathematical and informatics approaches to modeling biological systems in particular gene networks. Students are expected to have some experience with programming.

562 Statistics for Environmental Scientists (ECOL 562, ENST 562) (4). See ECOL 562 for description.

563 Statistical Analysis in Ecology and Evolution (ECOL 563, ENST 563) (4). Prerequisites, MATH 231 and STOR 151. Permission of the instructor for students lacking the prerequisites. Application of modern statistical analysis and data modeling in ecological and evolutionary research. Emphasis is on computer-intensive methods and model-based approaches. Familiarity with standard parametric statistics is assumed.

BME

BME 512: Biomedical Signal Processing (NCSU only?)
Fundamentals of continuous and discrete-time signal processing as applied to problems in biomedical instrumentation. Properties of biomedical signals and instruments. Descriptions of random noise and signal processes. Interactions between random biomedical signals and systems. Wiener filtering. Sampling theory. Discrete-time signal analysis. Applications of Z-transform and discrete Fourier transform. Digital filter design methods for biomedical instruments.

BME 590.006. Biosensors and bioanalytical measurements are critical tools in modern healthcare. They provide the information required to assist physicians in diagnosing disease. This course is focused on providing engineers with the background and understanding needed to design, build and interpret data from bioanalytical sensors. No previous biosensor or biological experience is required for this course. Topics include optical and electrochemical measurements, the operation of biological recognition elements, and design requirements for biochemical measurements.

Chem

070 First-Year Seminar: You Don't Have to Be a Rocket Scientist (3).
The goal of this seminar is to develop tools for extracting information from or finding flaws in news reports and popular science writing. Group work on such issues as biomass fuels, the hydrogen economy, and other alternative energy sources will develop an understanding of their economic and environmental impact.

Computer Science (*below includes only selected choices)

110 Introduction to Programming (3). Introduction to computer use. Approaches to problem solving; algorithms and their design; fundamental programming skills. Students can receive credit for only one of COMP 110, 116, or 121.

116 Introduction to Scientific Programming (3). Prerequisite, MATH 231. An introduction to programming for computationally oriented scientists. Fundamental programming skills, using MATLAB or Python. Problem analysis, algorithm design, plotting and visualizing data, with examples drawn from simple numerical and discrete problems. Students can receive credit for only one of COMP 110, 116, or 121.

401 Foundation of Programming (4). A first formal course in computer programming is required. Advanced programming: Program specifications, preconditions, postconditions, loop invariants. Linear data structures, searching, and sorting. Algorithm paradigms and analysis.

410 Data Structures (3). Prerequisite, COMP 401. The analysis of data structures and their associated algorithms. Abstract data types, lists, stacks, queues, trees, and graphs. Sorting, searching, hashing.

521 Files and Databases (3). Prerequisites, COMP 410 and 411 and MATH 381. Placement of data on secondary storage. File organization. Database history, practice, major models, system structure and design.

535 Introduction to Computer Security (3). Prerequisites, COMP 410 and MATH 381. Principles of securing the creation, storage, and transmission of data and ensuring its integrity, confidentiality and availability. Topics include access control, cryptography and cryptographic protocols, network security, and online privacy.

555 Bioalgorithms (3). Prerequisites, COMP 410 and MATH 381. Bioinformatics algorithms. Topics include DNA restriction mapping, finding regulatory motifs, genome rearrangements, sequence alignments, gene prediction, graph algorithms, DNA sequencing, protein sequencing, combinatorial pattern matching, approximate pattern matching, clustering and evolution, tree construction, Hidden Markov Models, randomized algorithms.

715 Visualization in the Sciences (MTSC 715, PHYS 715) (3). Computational visualization applied in the natural sciences. For both computer science and natural science students. Available techniques and their characteristics, based on human perception, using software visualization toolkits. Project course.

722 Data Mining (3). Prerequisites, COMP 550 and STOR 435. Data mining is the process of automatic discovery of patterns, changes, associations, and anomalies in massive databases. This course provides a survey of the main topics (including and not limited to classification, regression, clustering, association rules, feature selection, data cleaning, privacy, and security issues) and a wide spectrum of applications.

CEE

ENST 167 Advanced Functions of Temporal GIS (3). The course focuses on the development of advanced functions for field-based Temporal Geographical Information Systems (TGIS). These fields describe natural, epidemiological, economic, and social phenomena distributed across space and time. The course is organized around 4 main themes: constituents, mathematical framework, computer programs, and applications. Fall. Staff.

ENST 305, Data Analysis and Visualization of Social and Environmental Interactions (4). Prerequisites, ENST 201; MATH 231; or ECON 400. Principles of spatial and temporal data analysis are applied to issues of the role of society in producing environmental change. Methods include statistical analysis, model development and computer visualization. Three lecture hours and one lab hour a week. Spring. Staff.

ENST 468 (ENVR 468) Advanced Functions of Temporal GIS (3). Prerequisites math 233. Advanced functions of temporal geographical information systems (TGIS). These fields describe natural, epidemiological, economic and social phenomena distributed across space and time.

ENST 562 Statistics for Environmental Scientists.

ENST 563 Statistics for Ecology and Evolution.

ExSS

700 Applied Statistics and Research Methods in Exercise and Sport Science (3). Required preparation, undergraduate statistics course. Applied statistical analysis - interpretation of data from exercise and sport science. Emphasis: choosing method of analysis, using statistics software to run analyses. Major topics: experimental and nonexperimental research design, sampling, hypothesis testing, power calculation, t-tests, ANOVA, correlation, simple and multiple regression, and chi square.

705 Applied Statistics and Research Methods Laboratory (3). Required preparation, any undergraduate statistics course. Builds heavily upon material presented in EXSS 700. Planning, conducting, and reporting of research. Thesis writing and writing for publication. Problem-solving and practical experience in applied statistical analysis, interpretation, and presentation of data from the field of exercise and sport science.

GEOL

483 Geologic and Oceanographic Applications of Geographical Information Systems (MASC 483) (4). Required preparation, four GEOL courses or permission of the instructor. Focus is on applying GIS concepts and techniques to mining and petroleum geology, resource assessment, hydrogeology, coastal

and marine geology, physical oceanography, engineering geology, and a geologic perspective on land use. Three lecture and two laboratory hours a week.

520 Data Analysis in the Earth Sciences (3). Prerequisites, MATH 231 and 232. Required preparation, an introductory geology course numbered below 202, except first-year seminar, or permission of the instructor. Introduction to quantitative analysis in earth sciences: solid earth, atmospheres, oceans, geochemistry, and paleontology. Topics covered: univariate and multivariate statistics, testing, nonparametric methods, time series, spatial and cluster analysis, shapes.

Marine Sciences (MASC)

51 First-Year Seminar: Global Warming: Serious Threat or Hot Air? (3). Students will examine evidence that human activity has caused global warming, investigate scientists' ability to predict climate change, and discuss the political and social dimensions of global climate change.

561 Time Series and Spatial Data Analysis (3). Prerequisite, MATH 233. Permission of the instructor for students lacking the prerequisite. Three components: statistics and probability, time series analysis, and spatial data analysis. Harmonic analysis, nonparametric spectral estimation, filtering, objective analysis, empirical orthogonal functions.

Math

MATH 56H: Information and Coding. It is common to say that we are now living in the information age. What are the ways in which information is stored, transmitted, presented, and protected? What is information anyway? Topics for this seminar will be drawn from cryptography (secret writing throughout history, including Thomas Jefferson's cipher machine, the German Enigma machine, and security and privacy on the internet); image compression and processing (compact disks, MP3 and JPEG, transforms, error correction, noise removal); symbolic dynamics (encoding of symbol streams, like the genetic code, and associated dynamical systems and formal languages); and visualization (how can different kinds of information be vividly and usefully presented, combined, and compared?) These topics are mathematically accessible to anyone with a high school background and offer many possibilities for experimentation and theoretical exploration. Students will undertake individual or group projects using existing software for encoding and decoding messages, enhancing and compressing images, transforming and filtering signals, measuring properties of information sources, and so on. They will report on their work in writing and orally to the seminar. Discussions will be based on readings from a

course pack as well as Simon Singh's The Code Book (Doubleday, 1999), with investigations in probability, number theory, combinatorics, and information theory to provide theoretical foundations.

MATH 67: Mathematics of Climate Change. There is widespread agreement in the scientific community that the Earth is warming. But, do we know when critical benchmarks will be reached? Planning and policy-making demand predictions of future climate change and even specific climate events, but these predictions are based largely on complex mathematical models containing assumptions and estimations applied to chaotic dynamical systems. We must ask if it is even possible to make predictions about the future climate. Background on climate change will be covered in this seminar and extensive discussions will be held about what we know and what we do not know. The emphasis will then be on the issues surrounding predictability of climate events and changes and the limitations of mathematical models in relation to making predictions. This is an exciting scientific area where applied mathematics comes together with many other scientific areas in a political context that is of enormous importance to us all.

There is plenty of room for different viewpoints and deep thinking about how mathematics can contribute. Considerable time will be given to open discussions in class. There will be readings, and groups of students will make presentations on relevant topics as part of the course requirement. Further, each student will conduct a project and report on it at the end of the semester.

MATH 89: Networks: The Science of the Connected World. We live in a connected world, where the confluence of the different connections—social, political, financial, informational, technological, biological, behavioral, epidemiological—affects virtually every aspect of our lives. The study of networks provides a language for describing these connections and for attempting to describe the resulting impacts. Most people are familiar with the concept of a network in terms of hyperlinked web pages or online social networks; but networks are also useful for studying a wider variety of applications, with “nodes” representing actors of interest and connecting “edges” representing relationships. We will explore the roles of networks in public health, political activity, economic markets, workplace interactions, and internet search, among others. We will explore classical ideas in graph theory, fundamental concepts in social network analysis, and more recent developments in network science. We will also meet with guest speakers who are leading mathematical and social scientists studying networks.

406 Mathematical Methods in Biostatistics (1). Prerequisite, MATH 232. Special mathematical techniques in the theory and methods of biostatistics as related to the life sciences and public health. Includes brief review of calculus, selected topics from intermediate calculus, and introductory matrix theory for applications in biostatistics.

547 Linear Algebra for Applications (3). Prerequisite, MATH 233 or 283. Algebra of matrices with applications: determinants, solution of linear systems by Gaussian elimination, Gram-Schmidt procedure, eigenvalues. MATH 416 may not be taken for credit after credit has been granted for MATH 547.

Phys

PHYS 715 is a cross-list of COMP 715.

Psyc

53 First-Year Seminar: Talking about Numbers: Communicating Research Results to Others (3). This course introduces the many ways that research results are disseminated to the public in our everyday lives—through advertising and mass media, Internet, and research-based policy statements.

210 Statistical Principles of Psychological Research (3). Prerequisite, PSYC 101. Consideration of the methodological principles underlying psychological research, descriptive and inferential techniques, and the manner by which they may be employed to design psychological experiments and analyze behavioral data. Three lecture hours and one laboratory hour a week.

215 Statistical Principles of Psychological Research (B.S. Majors) (3). Prerequisites, MATH 231 and PSYC 101. Considers the methodological principles underlying psychological research, descriptive, and inferential techniques. This section is for students in the B.S. psychology program or for B.A. students with a calculus background and strong interest in quantitative psychology. Note: PSYC 215 will substitute for PSYC 210 as a prerequisite for other psychology courses.

532 Quantitative Psychology (3). Prerequisite, PSYC 210 or 215 or SOCI 252 or STOR 155. This course examines the science of quantitative psychology. Topics include the analysis of data, the design of questionnaires, and the assessment of psychological attributes, among others.

533 The General Linear Model in Psychology (3). Prerequisite, ECON 400 or PSYC 210 or 215 or SOCI 252 or STOR 155. Consideration of multiple regression and the general linear model in psychological research, including hypothesis testing, model formulation, and the analysis of observational and experimental data.

830 Statistical Methods in Psychology I (4). Required preparation, a course in introductory statistics. Data analysis, sampling, applied probability, elementary distribution theory, principles of statistical inference.

831 Statistical Methods in Psychology II (4). Prerequisite, PSYC 830. Statistical estimation and hypothesis testing for linear models (ANOVA, ANCOVA, regression analysis); statistical models in the design and analysis of experiments.

840 Computational Statistics (3). Prerequisite, PSYC 831. Permission of the instructor for students lacking the prerequisite. Current computational environments for data analysis and visualization are taught and used as a basis for understanding current (and creating new) methods of computational statistics and dynamic statistical graphics.

841 Introduction to Multivariate Techniques for the Behavioral Sciences (3). Prerequisite, PSYC 831. Permission of the instructor for students lacking the prerequisite. An introduction to linear regression and multivariate statistical techniques as employed in the behavioral sciences, with particular emphasis on analytic techniques and interpretation of results.

843 Factor Analysis (3). Prerequisite, PSYC 831. Permission of the instructor for students lacking the prerequisite. Advanced topics in factor analytic models, multivariate correlational models and analysis of covariance structures as applied in behavioral research.

844 Structural Equation Models with Latent Variables (3). Prerequisite, PSYC 831. Permission of the instructor for students lacking the prerequisite. Examination of a wide range of topics in covariance structure models, including their history, underlying theory, controversies and practical use with major computer packages.

845 Latent Curve Modeling (3). Prerequisite, PSYC 844. Permission of the instructor for students lacking the prerequisite. Latent curve modeling is a structural equations-based method for analyzing longitudinal data. Equal emphasis is placed on the statistical model and applications to real data.

846 Multilevel Modeling (3). Prerequisites, PSYC 830 and 831. This course demonstrates how multilevel models (or hierarchical linear models) can be used to appropriately analyze clustered data (i.e. persons within groups) and/or repeated measures data in psychological research.

853 Analysis of Frequency Tables in Behavioral Research (3). Prerequisite, PSYC 831. Permission of the instructor for students lacking the prerequisite. An introduction to the analysis of frequency data (including measures of association) and the use of log-linear models and logit models in the behavioral sciences.

854 Quantitative Research Synthesis (3). Prerequisite, PSYC 831. Permission of the instructor for students lacking the prerequisite. Survey of research synthesis including history, problem formulation, statistical concerns, describing and combining studies, combining p-values, testing for heterogeneity, accounting for moderator variables, fixed, mixed, and random effects models, publication bias.

STOR (*below includes only selected choices)

STOR 053 Networks: Degrees of Separation and Other Phenomena Relating to Connected Systems (3). Networks, mathematical structures that are composed of nodes and a set of lines joining the nodes, are used to model a wide variety of familiar systems: distribution networks such as electric power grids, anatomical networks such as neural systems, communication networks such as the world-wide web, and social networks representing relationships between cultural groups. These networks have distinct properties that help answer questions about the underlying system: how susceptible is a power grid to breakdown? how fast can a computer virus spread? how connected are the members of different corporate boards? Questions of this type, some suggested by class members, will be posed and modeled by networks.

STOR 072 Using Computers to Unlock the Genetic Code (3). First Year Seminar.An introduction to DNA - its structure, function, and importance. Includes topics from computational, organizational and statistical tools for unlocking the secrets of life. Provan.

STOR 112 Decision Models for Business (3). Prerequisite, Math 110 or exemption. An Introduction to the basic quantitative models of business with linear and non-linear functions of single and multiple variables. Linear and non-linear optimization models and decision models under uncertainty will be covered. Fall and Spring. Staff.

STOR 113 Decision Models for Economics (3). Prerequisite, Math 110 or exemption. An Introduction to multi-variable quantitative models in economics. Mathematical techniques for formulating and solving optimization and equilibrium problems will be developed, including elementary models under uncertainty. Fall and Spring. Staff.

STOR 151 Basic Concepts of Statistics and data analysis (3). Prerequisite, Math 110 or exemption. Elementary introduction to statistical reasoning, including sampling, elementary probability, statistical inference, and data analysis. STOR 151 may not be taken for credit by students who have credit for ECON 400 or PSYC 210. Fall, Spring, Summer. Staff.

STOR 155 Introductory Statistics (3). Prerequisite, Math 110 or exemption. Data analysis: Visual measures, histogram, scatterplot, etc; Quantitative measures, central tendency and dispersion, correlation;Simple linear regression; Exploratory data analysis, outliers, leverage. Sampling, estimation of parameters; Probability: Basic rules; Conditional probability, independence; Random variables, distributions, expected values, variance; Binomial and normal distributions. Sampling Distributions: Distribution of sample mean; Central limit theorem. Inference: Hypothesis testing; Confidence intervals for population mean, binomial proportion; Regression parameters.

STOR 215 Introduction to the Decision Sciences (3). Prerequisite, Math 110 or exemption. Introduction to basic concepts and techniques of decision-making and information management in business, economics, social and physical science. Topics include discrete optimization, discrete probability, networks, decision trees, games, Markov chains. Fall. Staff. General College/B.A. -level Mathematical Science Perspective.

STOR 305 Mathematical Models for Decision Making (3). Prerequisite: STOR 155. The use of mathematics to describe and analyze large-scale decision problems. Situations involving the allocation of resources, making decisions in a competitive environment, and dealing with uncertainty are modeled and solved using suitable software packages. Fall. Wagner.

STOR 358 SAMPLE SURVEY METHODOLOGY (BIOS 664) (3). Prerequisite: STOR 456 or equivalent. Principles and methods associated with survey sampling, including simple random sampling, stratified sampling and cluster sampling. Questionnaire design, problems of nonresponse, sources of nonsampling errors. Design, execution, and analysis of an actual survey. Spring. Kalsbeek.

STOR 415 Deterministic Models in Operations Research (3). Prerequisite: MATH 547. Linear, integer, nonlinear and dynamic programming, classical optimization problems, network theory. Spring. Pataki, Provan, Lu.

STOR 435 Introduction to Probability (MATH 535) (3). Prerequisites: Working knowledge of MATH 231-233. Introduction to the mathematical theory of probability. Probability models for random experiments. Basic properties of probability measures. Conditional probability and independence. Discrete random variables: hypergeometric, binomial, geometric, negative binomial, and Poisson. Continuous random variables: uniform, exponential, Gaussian, Cauchy and gamma. Jointly distributed random variables. Definition and basic properties of expectations, variances, covariances and correlations. Basic inequalities for probabilities and expectations. Laws of large numbers and the central limit theorem. Fall and Spring. Budhiraja, Nobel.

STOR 445 Stochastic Models in Operations Research (3). Prerequisite: STOR 435. Introduction to Markov chains, Poisson process, continuous-time Markov chains, renewal theory. Applications to queueing systems inventory, and reliability, with emphasis on systems modeling, design, and control. Fall. Kulkarni, Ziya.

STOR 455 STATISTICAL METHODS I (3). Prerequisite: STOR 155 or equivalent. Some familiarity with matrix algebra recommended, but not required. This course presents regression analysis and related techniques, and is recommended for students throughout the natural and social sciences who are interested in applying regression analysis in their research and/or understanding the statistical concepts underlying the methodology. The topics include simple and multiple linear regression, matrix representation of the regression model, statistical inferences for regression model, diagnostics and remedies for multicollinearity, outlier and influential cases, polynomial regression and interaction regression models, model selection, weighted least square procedure for unequal error variances, and ANOVA model and test. Statistical software SAS will be used throughout the course to demonstrate how to apply the techniques on real data. The main purposes of this courses is to let students know how to use regression methods properly in data analysis and lay the foundation for more advanced studies in statistics. Fall. Hannig.

STOR 456 Statistical Methods II (3). Prerequisite: STOR 455. The focus of this course is on analysis of time series data, that is, data recorded in time. Such data arise in a wide range of areas including Environmental Sciences, Economics, Business and Finance, Actuarial Sciences, Social Sciences. The topics of the course include estimation and elimination of trend and seasonal components, stationary time series, ARMA models, spectral analysis, modeling and forecasting of time series. Some statistical software will be used throughout the course to demonstrate how to apply the techniques on real time series data. Spring. Shen, Pipiras, Smith.

STOR 555 Mathematical Statistics (3). Prerequisite: STOR 435 or equivalent. Derivation and analysis of point estimators using decision theory, including the methods of Bayes and minimax estimation, maximum likelihood, method of moments, and unbiased estimation. Confidence intervals. Hypothesis testing, including Bayesian methods, multiple hypotheses, the Neyman-Pearson Lemma, simple and composite hypotheses, likelihood ratio tests, Type I and Type II errors, power calculations. Uses of and relationships between the families of standard probability models, including the Normal, Gamma, Chi-Squared, Student's T, Uniform, Beta, Binomial, Negative Binomial, Poisson, Hypergeometric, and Cauchy distributions, as well as the Poisson Process and the Bernoulli Process. Fall only. Carlstein, Kelly.

STOR 664 Applied Statistics I. Prerequisite: STOR 555 or permission of the instructor. Strong background in multivariate calculus and linear algebra is required for this course. This course presents in-depth regression analysis and related techniques. Both theoretical development and computational implementation of the techniques will be presented. The topics include simple and multiple linear regression; matrix representation of the regression model and solution; the geometric approach to least squares theory; the ANOVA table; confidence and prediction intervals; hypothesis tests; diagnostics for influential observations and model assumptions such as normality; diagnostics for model selection (multicollinearity, variable selection, transformations such as log, Box-Cox, etc.). Selected additional topics depending on time and instructor: weighted and generalized least squares, nonlinear regression, and analysis of designed experiments (one-way and two-way analysis of variance, factorial designs). Computing environments: R, SAS, MATLAB (varies depending on instructor). Fall. Carlstein, Liu, Smith. (3)

STOR 665 Applied Statistics II. Prerequisite: STOR 664 or permission of the instructor. Analysis of variance (ANOVA), including nested and crossed models, and multiple comparisons. GLM basics, including exponential families, link functions, likelihood, quasi-likelihood, conditional likelihood, iterative reweighted least squares, EM algorithm, and diagnostic techniques. Linear mixed model (LMM) basics including variance components, prediction of random effects; various estimation techniques based on ANOVA, maximum likelihood and restricted maximum likelihood. Selected additional topics depending on time and instructor: generalized estimating equations (GEE), generalized LMM, longitudinal data and various nonlinear models. Spring. Shen. (3)

STOR 754 Time Series and Multivariate Analysis. Prerequisite: STOR 435. Time Series: Exploratory and graphical analysis; Time domain analysis and ARMA models; Fourier analysis: FFT, periodogram, smoothing; State space analysis: Kalman filter, dynamic models. Multivariate: Principal components, canonical correlation; Classification, clustering; Dimension reduction: projection pursuit, alternating conditional sliced inverse regression. Spring. Leadbetter, Smith, Sen. (3)

STOR 757 Bayesian Statistics and Generalized Linear Models. Corequisites: STOR 664 and 655, or permission of the instructor. Bayes factors; Empirical Bayes, formulation, Stein effect; Classical: EM, Laird-Ware; Hierarchical: prior, MCMC. GLM specific models: Binomial regression, polytomous regression, Cox proportional hazard, log linear. Hannig, Smith. (3)

Data Studies current course inventory: Social Sciences
June 2013
Thomas Carsey
Odum Institute

Methodology:

ConnectCarolina. All courses that dealt with research methods and data, whether quantitative or qualitative in nature, were included in order to be as comprehensive as possible. I know there have been “Special Topics” courses offered in Political Science that focus on methods which are not included in the list below – I suspect that is true for other departments as well.

Also missing from the list below are the approximately 70 workshops that are offered by the Odum Institute every year.

African and Afro American Studies
None

American Studies
275 – Documenting Communities (3). Fulfills: SS; CI; E4. Covers the definition and documentation of communities within North Carolina through research, study, and field work of communities. Each student produces a documentary on a specific community.

275H – Documenting Communities (3). Fulfills: SS; CI; E4. Covers the definition and documentation of communities within North Carolina through research, study, and field work of communities. Each student produces a documentary on a specific community.

701 – Interdisciplinary Research Methods (3). This course will focus on techniques of American studies investigation. Various faculty members will make presentations highlighting approaches including Southern studies, American Indian studies, Material Culture studies, and new media.

840 – Digital Humanities/Digital American Studies (3). This course, explores the application of digital technologies to the materials, questions, and practices of humanities scholarship, particularly as related to enduring topics in American Studies scholarship and community engagement. Students will work on group digital history projects in collaboration with local cultural heritage organizations.

850 – Digital Humanities Practicum (3). This practicum blends graduate seminar discussions with hands-on training in the digital humanities. Students will work in the Digital Innovation Lab, contributing to real-life projects while developing their own professional development goals. Students will emerge with a deeper understanding of and experience with digital humanities approaches, practices, and issues.

948 – Research in Native American History (3). This course introduces graduate students to research methods in Native American history, including the methodology of ethnohistory and the techniques of compiling a source base, taking notes, and outlining.

Anthropology
416 – Bioarchaeology (3). Fulfills: SS. The study of human skeletal remains from archaeological contexts. The collection and interpretation of quantitative and qualitative data is emphasized to assess the relationship between past biology, environment, culture, and behavior.

415L – Zooarchaeology Lab (1). Corequisite, ANTH 415. Required preparation, an archaeological course or permission of instructor. Examination of identification techniques, quantitative methods, and interpretive frameworks used to analyze animal remains recovered from archaeological sites.

419 – Anthropological Application of GIS (3). Permission of the instructor. GIS experience required. This course explores applying GIS science technologies to anthropological problems. Students will learn GIS skills and apply them using spatial data.

455 – Ethnohistory (3). Fulfills: HS. Integration of data from ethnographic and archaeological research with pertinent historic information. Familiarization with a wide range of sources for ethnohistoric data and practice in obtaining and evaluating information. Pertinent theoretical concepts will be explored.

723 – Seminar in Anthropological Linguistics (3). Selected topics from general linguistics and sociolinguistics, special emphasis on methods and problems involved in analysis and description of semantic structure of language and its relation to the rest of culture.

724 – Anthropology and Cybernetics (3). Examination of systems theory, or cybernetics; evaluation of previous applications of cybernetic models in anthropology; and original analysis of anthropological data in these terms by students.

725 – Quantitative Methods in Anthropology (3). Survey of standardized data-gathering techniques, problems in research design, and methods of quantitative analysis encountered in anthropological research.

726 – Quantitative Methods in Archaeology (3). Introduction to quantitative and computer methods in archaeology. The course stresses exploratory data analysis and graphical pattern recognition techniques.

729 – Research Strategies in Archaeology (3). This seminar develops student's skills in crafting research designs, proposals, and presentations. Examples and readings focus on archaeology and bioarchaeology but the skills covered are widely applicable.

City and Regional Planning

491 – Introduction to GIS (3). Prerequisite, GEOG 370. Permission of the instructor for students lacking the prerequisite. Stresses the spatial analysis and modeling capabilities of organizing data within a geographic information system. (GISci)

591 – Applied Issues in Geographic Information Systems. Prerequisite, GEOG 370 or 491. Applied issues in the use of geographic information systems in terrain analysis, medical geography, biophysical analysis, and population geography.

701 – Research Methods (1-6). Course combines material learned in other courses (theory/philosophy, methods, and their substantive area of interest). Familiarizes students with the skills necessary to conduct research and critically review and understand evaluation reports.

720 – Planning Methods (3). Permission of the instructor for undergraduates. Accessing information from conventional and electronic sources, spatial data acquisition, analysis and mapping. Inferential statistics through multiple regression. Microcomputer laboratory.

721 – Advanced Planning Methods (1-6). Permission of the instructor for undergraduates. More in-depth treatment of topics covered in PLAN 720. Particular emphasis on techniques of multiple regression analysis, forecasting, categorical data analysis, and spatial data analysis.

739 – Transportation Planning Models (3). Permission of the instructor for undergraduates. The transportation planning process; data collection, trip generation, modal choice, trip distribution and assignment. Social, economic, and environmental impacts of transportation. Innovative modeling techniques.

741 – Land Use and Environmental Planning (3). Methods of land use planmaking. Use of GIS and spreadsheets to analyze land suitability and spatial needs. Preparation of land classification plans, land use design plans, and development management programs.

745 – Development Impact Assessment (3). Methods for data management and predictive analysis of the environmental, transportation, and other infrastructure; fiscal and social impacts of land development projects. Impact mitigation measures are also examined.

760 – Real Estate Investment and Affordable Housing (3). Fundamentals and techniques of real estate investment analysis, including cases and computer modeling; applications of the public interest in private investment decisions; tax and other public policies influencing real estate investments; and affordable housing.

771 – Development Planning Techniques (3). Intermediate and advanced techniques for analyzing the development of local and regional economies. Social accounts, indicator construction, regional input-output models, economic and fiscal impact analysis, labor market analysis, and regional economic forecasting techniques.

802 – Advanced Seminar in Research Design: Data, Methods, and Evaluation (3). Three main objectives: to deepen students' understanding of important issues and topics in the design of empirical research, to further develop students' ability to critically evaluate research designs and policy-related products, and to aid in developing a research paper, dissertation, or other product.

Communication Studies

610 – Reading Quantitative Research in Communication Studies (3). Permission of the instructor for nonmajors. Review of the basics of quantitative research (e.g., scientific method, modes of data collection, instrument development, data analysis techniques) with the goal of gaining skill in reading published articles in communication studies journals.

664/695 – Field Methods (3). Recommended preparation, COMM 562 or 841. A bridge course designed to offer graduate students and advanced undergraduates a practicum in fieldwork methods and performance ethnography.

723 – Research in Organizational Communication (3). Explores theoretical, methodological, and practical issues encountered in ethnographic, case study, and field research on communication phenomena in organizations.

798 – Topics in Research Methods (3). Advanced study of selected topics in research methods. Topics vary.

841 – Performance Ethnography (3). This seminar focuses on methods of ethnography and fieldwork ethics. Performance as theory and practice informs methodological inquiries as well as the analysis of specific ethnographic texts and case studies.

851 – Research Methods in Media and Cultural Studies (3). Graduate standing required. Introduction to the issues, methods, and materials of research in media and cultural studies.

Economics

400 – Elementary Statistics (3). Fulfills: QI. Comprehensive introduction to statistics, including descriptive statistics and statistical graphics, probability theory, distributions, parameter estimation, hypothesis testing, simple and multiple regression, and use of powerful statistical estimation software.

400H – Elementary Statistics (3). Fulfills: QI. Comprehensive introduction to statistics, including descriptive statistics and statistical graphics, probability theory, distributions, parameter estimation, hypothesis testing, simple and multiple regression, and use of powerful statistical estimation software.

511 – Game Theory in Economics (3). Prerequisites, ECON 410 and MATH 233. Permission of the instructor for students lacking the prerequisites. Topics in noncooperative and cooperative game theory are covered, along with a selection of applications to economics in areas such as industrial organization, international trade, public finance, and general equilibrium.

511H – Game Theory in Economics (3). Prerequisites, ECON 410 and MATH 233. Permission of the instructor for students lacking the prerequisites. Topics in noncooperative and cooperative game theory are covered, along with a selection of applications to economics in areas such as industrial organization, international trade, public finance, and general equilibrium.

570 – Economic Applications of Statistical Analysis (3). Fulfills: SS; E6; QI. Prerequisite, ECON 400. Statistical methods in the construction, estimation, testing, and application of linear economic models; computer programs and interpretation of their output in empirical analysis of common economic theories.

570H – Economic Applications of Statistical Analysis (3). Fulfills: SS; E6; QI. Prerequisite, ECON 400. Statistical methods in the construction, estimation, testing, and application of linear economic models; computer programs and interpretation of their output in empirical analysis of common economic theories.

575 – Econometric Topics: Applied Time Series Analysis and Forecasting (3). Fulfills: SS; QI. Prerequisites, ECON 400, 410, 420, and 570. Permission of the instructor for students lacking the prerequisites. Econometric techniques for time series data. Topics include ARMA models, forecasting, nonstationarity, conditional heteroskedasticity, and multiple equation models.

700 – Basic Quantitative Techniques (3). Topics from linear algebra, calculus, linear and nonlinear programming, and the theory of difference and differential equations with applications to economics.

770 – Introduction to Econometric Theory (3). Probability theory, expectation, conditional expectation, modes of convergence, limit and interchange theorems, and the asymptotics of maximum likelihood, generalized method of moments and efficient method of moments.

771 – Econometrics (3). Prerequisite, ECON 770. One semester coverage of basic econometrics. Topics include: regression under ideal and nonideal conditions; special models, including simultaneous equations models; and applications and econometric computer programs.

810 – Game Theory I (3). Prerequisite, ECON 710 and 711. Permission of the instructor for students lacking the prerequisites. Noncooperative games in strategic and extensive form, with perfect and imperfect information. Other topics from: information economics, mechanism design, auctions, repeated games, bargaining, bounded rationality, learning, evolutionary games, cooperative games.

811 – Game Theory II (3). Prerequisite, ECON 810. Permission of the instructor for students lacking the prerequisite. This course is a continuation of ECON 810. Topics covered will be chosen from those listed, but not covered in ECON 810.

850 – Health Economics (3). Prerequisites, ECON 710 and 771. Permission of the instructor for students lacking the prerequisites. Measurement and modeling of the demand for medical care, the demand for and supply of health insurance, and the incorporation of health, medical care, and health insurance in determining both short and long run labor supply.

851 – Health Economics for Developing Countries (3). Prerequisites, ECON 710 and 771. Permission of the instructor for students lacking the prerequisites. Major topics are: how health and development are related, the demand for health services, cost-benefit and cost-effectiveness analysis, and methods for financing health care in developing, resource-constrained nations.

855 – Economics and Population (3). Prerequisite, graduate standing in economics or permission of the instructor. Analysis of economic-demographic interrelationships including: population and economic development; population, environmental decay, and zero population growth; models of fertility, migration, and spatial organization; population policy. (Not regularly offered.)

870 – Advanced Econometrics (3). Prerequisites, ECON 770, 771, and MATH 547. ECON 870 constitutes a one-semester treatment of the fundamental theory of econometrics. Topics covered include asymptotic distribution theory, linear and nonlinear models, specification testing techniques, and simultaneous equations models.

871 – Time Series Econometrics (3). Prerequisite, ECON 870. Covers stationary univariate and multivariate time series models, spectral analysis methods, nonstationary models with time trends, unit roots and cointegration, and special topics such as conditional volatility, the Kalman filter, and changes of regime.

872 – Nonlinear Econometric Methods (3). Prerequisite, ECON 870. Density estimation, nonparametric regression, neural nets, nonlinear regression, generalized method of moments, seminonparametric time series, estimating stochastic differential equations and nonlinear latent variables.

873 – Microeconomics (3). Prerequisite, Economics 870. Limited dependent variable models such as binary outcome models, multinomial outcome models, and censored and truncated outcome models. Count data models. Duration models. Panel data analysis.

876 – Advanced Topics in Empirical Finance (3). Prerequisite, ECON 871. This course will cover a selected list of current empirical research topics in finance and related econometric methods.

877 – Foundations for Continuous Time Asset Pricing (3). Prerequisites, STOR 634 and 635. This course introduces students to mathematical foundations and economic interpretation of the main probabilistic tools (stochasatic calculus, martingale methods) in continuous time finance.

971 – Research in Econometrics (3). The course introduces students to theoretical and applied research topics in econometrics. May be repeated for credit.

Geography

370 – Introduction to Geographic Information (3). A survey of geographic data sources including maps, photos, digital images, Census information, and others. Emphasis is on appropriate uses, limitations, and skilled interpretation in physical and human geography applications. (Core)

391 – Quantitative Methods in Geography (3). This course provides an introduction to the application of statistical methods to geographic problems and to statistical packages in their solution. Attention given to spatial data analysis and sampling methods.

392 – Research Methods in Geography (3). Fulfills: SS. Introduces geographic research methods and develops skills to conduct independent research. Themes include spatial analysis, knowledge production, methodology, theory and evidence, and principles of informed argument. Students gain experience with multiple methods applicable to the study of diverse topics.

410 – Modeling of Environmental Systems (3). Fulfills: QI. Uses systems theory and computer models to understand ecosystem energy and matter flows, such as energy flow in food webs, terrestrial ecosystem evapotranspiration and productivity, related to climate, vegetation, soils, and hydrology across a range of spatial and temporal scales.

410 – Modeling of Environmental Sciences (3). Prerequisite, GEOG 110. Use of systems theory and computer modeling to understand general issues in climate, vegetation, geomorphology, soils, and hydrology such as crossing time and space scales and linear and dynamical systems. No laboratory. (GISci)

416 - Applied Climatology: The Impacts of Climate and Weather on Environmental and Social Systems (3) Fulfills: PL. Research Intensive. Applied climatology involves the interdisciplinary application of climate data and techniques to solve a wide range of societal and environmental problems. This projects-based course investigates how climate impacts a range of sectors, including water resources, urban environments, ecosystems, and human health.

416 – Applied Climatology (3). Prerequisite, GEOG 110 or 111. Permission of the instructor for students lacking the prerequisite. An investigation of the ways climatic information and techniques can be applied

to environmental and societal problems, such as water resources, urban environments, vegetation, and human health. (EES)

419 – Field Methods in Physical Geography (3). Fulfills: E4. Involves evaluation of landscapes by examining nature and biophysical elements influencing landscape form and function. Course emphasizes data collection, analysis, and interpretation using GIS and field methods. (EES)

491 – Introduction to GIS (3). Prerequisite, GEOG 370. Permission of the instructor for students lacking the prerequisite. Stresses the spatial analysis and modeling capabilities of organizing data within a geographic information system. (GISci)

541 – GIS in Public Health (3). Explores theory and application of geographic information systems (GIS) for public health. The course includes an overview of the principles of GIS in public health and practical experience in its use. (GISci)

577 – Advanced Remote Sensing (3). Prerequisite, GEOG 370 or 477. Acquisition, processing, and analysis of satellite digital data for the mapping and characterization of land cover types. (GISci)

591 – Applied Issues in Geographic Information Systems (3). Prerequisite, GEOG 370, 491, or equivalent. Applied issues in the use of geographic information systems in terrain analysis, medical geography, biophysical analysis, and population geography. (GISci)

592 – Geographic Information Science Programming (3). Prerequisite, GEOG 370 or 491. This course will teach students the elements of GISci software development using major GIS platforms. Students will modularly build a series of applications through the term, culminating in an integrated GIS applications program.

593 – Geographic Information Science Programming (3). Prerequisite, GEOG 370 or 491. This course will teach students the elements of GISci software development using major GIS platforms. Students will modularly build a series of applications through the term, culminating in an integrated GIS applications program.

594 – Global Positioning Systems and Applications (3). Prerequisite, GEOG 370. Global Positioning Systems (GPS) fundamental theory, application design, post processing, integration of GPS data into GIS and GPS application examples (such as public health, business, etc.) will be introduced.

595 – Ecological Modeling (3). Prerequisite, BIOL 561 or STOR 455. Permission of the instructor for students lacking the prerequisite. This course focuses on modeling the terrestrial forest ecosystems processes, including population dynamics, energy, water, nutrients, and carbon flow through the ecosystem. (GISci)

597 – Ecological Modeling (3). Prerequisite, BIOL 561 or STOR 455. Permission of the instructor for students lacking the prerequisite. This course focuses on modeling the terrestrial forest ecosystems processes, including population dynamics, energy, water, nutrients, and carbon flow through the ecosystem. (GISci)

703 – Geographic Research Design (3). Introduction to the theory and practice of geographic research. The range of methods available for problem identification and solution are considered through development of specific research proposals.

705 – Advanced Quantitative Methods in Geography (3). Application of selected multivariate statistical techniques to the analysis of geographic phenomena and problems.

790 – Spatial Analysis and Computer Models (3). This course introduces students to spatial analysis techniques involving points, lines, areas, surfaces, and non-metric spaces, as well as programming basic geographic models on microcomputers.

History

670 – Introduction to Oral History (3). Fulfills: HS; CI. Introduces students to the uses of interviews in historical research. Questions of ethics, interpretation, and the construction of memory will be explored, and interviewing skills will be developed through field work.

746 – History and the Social Sciences (3). The relationship of the social sciences to history, logic of inquiry, use of quantitative methods, and introduction to the computer.

924 – Seminar in Modern European History (3). This writing seminar explores the process of working with primary sources, creating a narrative, and shaping an interpretation based on examples from the last two centuries of European history.

948 – Research in Native American History (3). This course introduces graduate students to research methods in Native American history, including the methodology of ethnohistory and the techniques of compiling a source base, taking notes, and outlining.

Linguistics

200 – Phonology (3). Prerequisite, LING 101. Description and analysis of sound patterns from languages around the world. Introduction to formal phonological models, written argumentation, and hypothesis testing.

301 – Language and Computers (3). Prerequisite, LING 101. Uses simple linguistic problems to introduce students to the use of programming languages especially suited to analyze and process natural language on the computer. No prior programming knowledge is presupposed.

422 – Research Methods in Phonetics and Laboratory Phonology (3). Prerequisite, LING 200, 520, 523, or SPHS 540. Focuses on the practical skills required to carry out basic experiments in speech production or perception. Includes training in a general-purpose programming language (such as Perl) for automating repetitive tasks, experiment-control software, audio stimulus manufacture and editing, palatography, aerodynamic measurements, and other laboratory techniques relevant to student interests.

522 – Experimental Phonetics and Laboratory Phonology (3). Prerequisites, LING 520, and 200 or 523. This course relates linguistic theory to experimental findings. Students design and carry out experiments to test theoretical issues of current theoretical importance.

540 – Mathematical Linguistics (3). Fulfills: QI. Introduction to topics in logic, set theory, and modern algebra with emphasis on linguistic application. Automata theory and the formal theory of grammar with special reference to transformational grammars. No previous mathematics assumed.

573 – Linguistic Field Methods I (3). Analysis and description of a language unknown to the class from data solicited from a native-speaker consultant.

574 – Linguistic Field Methods II (3). Analysis and description of a language unknown to the class from data solicited from a native-speaker consultant.

Political Science

181 Quantitative Research in Political Science (3). An introduction to 1) conceptual foundations of scientific study of politics, 2) research design, 3) descriptive statistics, and 4) inferential statistics. To accomplish these goals, the course employs class lectures, readings, and problem set assignments.

209 – Analyzing Public Opinion (3). Fulfills: SS; QI. A study of forces affecting public opinion and its expression in various political activities. Emphasis on gathering and analyzing opinion data. Course may be taught in the computer classroom.

209H – Analyzing Public Opinion (3). Fulfills: SS; QI. A study of forces affecting public opinion and its expression in various political activities. Emphasis on gathering and analyzing opinion data. Course may be taught in the computer classroom.

285 – Research Methods and Experiments (3). Fulfills: SS; QI. This course is designed to provide students with a thorough understanding of the various quantitative research methods available to researchers in social science. No prerequisite is required. The course will focus on experimental methods in political science, including laboratory experiments, field experiments, and survey experiments.

288 – Strategy and Politics (3). Fulfills: SS; QI. Offers an introduction to positive political theory, the application of rational choice analysis (or economic models) to the study of political phenomena. Topics include social choice theory, legislative voting, problems of cooperation and collective action, and public choice theory. Encourages students to think about politics from a critical vantage point.

289 – Strategy and International Relations (3). Fulfills: SS; GL. Introduction to the study of strategic decision making in international relations, with an emphasis on the application of basic game theoretic models. Incorporates in-class simulations of international relations scenarios.

384 – Introduction to Philosophy, Politics, and Economics (3). Fulfills: PH. Permission of the department required; one course in economics strongly recommended. This interdisciplinary gateway course provides an introduction to subjects and quantitative techniques used to analyze problems in philosophy, political science, and economics.

488 – Advanced Game Theory (3). Fulfills: SS; QI. Prerequisite, POLI 288 or 289. Increasingly, political and social scientists are using game theory to analyze strategic interactions across different settings. This course aims to give students a deep technical understanding of the most relevant concepts of game theory and how these concepts have been applied to the study of political and economic phenomena.

725 – Public Administrations Analysis and Evaluation II (3). Prerequisite PUBA 719. Second course in a two-course sequence introducing students to applied research design, data collection, data management, data analysis, and analytical reporting to allow students to conduct original research, be informed consumers of other research, and ultimately improve public program planning and evaluation decisions.

754 – Introduction to Mathematical International Relations (3). Surveys research in mathematical models of international decision making, bargaining, systemic change, arms races, coalitions, and perception. Philosophic and historical considerations about this field are also discussed.

780 – Scope and Methods of Political Research (3). Permission of the instructor. A discussion of the theory and process of political analysis, including philosophy of science, research design, the methods of drawing causal inferences, and of generating data.

783 – Statistics (3). Elementary descriptive statistics and basic principles of statistical inference including estimation and tests of hypotheses.

784 – Intermediate Statistics (3). This course extends the coverage of POLI 281. Topics to be covered include analysis of variance, multiple and partials correlation, and multiple regression.

785 – Introduction to Structural Equation Models (3). Prerequisite, POLI 784. Introduces structural equation models with observed variables and econometric estimation methods. Some attention to models with unobserved variables and LISREL-type analyses.

786 – Time Series Analysis of Political Data (3). Prerequisite, POLI 784. Permission of the instructor for students lacking the prerequisite. Discusses the problems that arise when regression methodologies are applied to time series and pooled time series data.

787 – Maximum Likelihood Methods (3). Prerequisites, POLI 783 and 784. Introduction to maximum likelihood estimation with applications to political science. Topics include discrete choice analysis, censored and truncated variables, event history analysis, sample selection models, and multilevel inference.

788 – Statistics and Data Analysis for Political Science and Policy Research (3). This course focuses on the application of statistical analysis to quantitative data in order to study theoretically and substantively interesting questions about politics and policy.

789 – Game Theory (3). This class provides graduate students with an introduction to game theoretic modeling, focusing on noncooperative game theory. Topics covered include normal form games, extensive-form games, and games of incomplete information.

790 – Positive Political Theory (3). This seminar surveys applications of rational choice models across the subfields of political science. It also considers critiques of national choice approaches and alternative theoretical approaches to modeling human behavior.

880 – Design and Analysis of Experiments and Surveys (3). Prerequisites, POLI 780 and 783. Introduction to the use of experimental and survey research methods in political science. Topics include: factorial designs, repeated measures design, ANOVA, sampling theory, survey errors and costs, and questionnaire design.

Psychology
Already covered in the inventory done in the Natural and Mathematical Sciences

Public Policy
460 – Quantitative Analysis for Public Policy (4). Fulfills: QI. Prerequisite, STOR 155. Application of statistical techniques, including regression analysis, in public policy program evaluation, research design, and data collection and management.

681 – Research Design for Public Policy (3). Pre- or corequisite, PLCY 460. Students will explore the scientific method as applied to policy research. They will formulate testable policy research questions, become familiar with methods for conducting policy research, and learn to think critically about causal inference.

700 – Mathematical Preparation for Public Policy and Economics (3). An intensive preparation course in mathematical and statistical analysis for public policy and economics. Reviews and introduces topics in linear algebra, calculus, optimization and mathematical statistics, and prepares students for PLCY 788 and PLCY 789. Also serves as a prerequisite for HPM 881, which satisfies one methods requirement in the Ph.D. program.

801 – Design of Policy-Oriented Research (3). An intensive preparation course in mathematical and statistical analysis for public policy and economics. Reviews and introduces topics in linear algebra, calculus, optimization and mathematical statistics, and prepares students for PLCY 788 and PLCY 789. Also serves as a prerequisite for HPM 881, which satisfies one methods requirement in the Ph.D. program.

802 – Advanced Seminar in Research Design: Data, Methods, and Evaluation (3). Three main objectives: to deepen students' understanding of important issues and topics in the design of empirical research, to further develop students' ability to critically evaluate research designs and policy-related products, and to aid in developing a research paper, dissertation, or other product.

882 – Advanced Panel Data Methodology for Public Policy (3). Three main objectives: to deepen students' understanding of important issues and topics in the design of empirical research, to further develop students' ability to critically evaluate research designs and policy-related products, and to aid in developing a research paper, dissertation, or other product.

Sociology

251 – Measurement and Data Collection (3). Required of sociology majors. Methods of data collection, with attention to problem selection, sources of information, choice of methods, and research design. Operationalization and measurement; sampling, construction of questionnaires, and interviewing; observation techniques; experimentation.

252 – Data Analysis in Social Research (3). Fulfills: QI. Prerequisite, SOCI 251. Required of sociology majors. Methods of data analysis: descriptive and inferential statistics and multivariate analysis to permit causal inference. Attention to problems of validity and reliability and to index construction.

620 – Aging and Cohort Analysis in Social and Epidemiologic Research: Models, Methods, and Innovations (3). Required preparation, basic statistics courses. This seminar introduces guidelines for conducting aging and cohort analysis in social and epidemiologic research in which time and change are concerns. Uses three common research designs with an emphasis on new analytic models and methods.

707 – Measurement and Data Collection (4). Provides an introduction to measurement theory and a review of various methods of data-gathering. Gaining experience with a variety of techniques of measurement and preparing a pretested research proposal are required for all students.

708 – Statistics for Sociologists (4). Provides an introduction to probability theory, descriptive statistics, inferential statistics, and the algebra of expectations. Emphasis is on elements useful to research sociologists, including bivariate regression and correlation.

709 – Linear Regression Models (4). The course presents regression analysis and related techniques. The major topics are the assumptions of the regression model, dummy variables and interaction terms, outlier diagnostics, multicollinearity, specification error, heteroscedasticity and autocorrelation. The final section introduces path analysis, recursive models, and nonrecursive systems.

711 – Analysis of Categorical Data (3). Permission of the instructor. Introduction to techniques and programs for analyzing categorical variables and nonlinear models. Special attention is given to decomposition of complex contingency tables, discriminant function analysis, Markov chains, and nonmetric multidimensional scaling.

715 – Seminar on Social Networks (3). Permission of the instructor. Theoretical and substantive issues in social network analysis. Focus is on models of social structure.

717 – Structural Equations with Latent Variables (3). Prerequisite, SOCI 708. Permission of the instructor. This course examines models sometimes referred to as LISREL models. Topics include path analysis, confirmatory factor analysis, measurement error, model identification, nonrecursive models, and multiple indicators.

718 – Longitudinal and Multilevel Data Analysis (3). Prerequisite, SOCI 709 or 711. This course provides an introduction to event history analysis or survival analysis, random effects and fixed effects models for longitudinal data, multilevel models for linear and discrete multilevel data, and growth curve models.

753 – Experimental Design in Sociology (3). Permission of the instructor. Statistical aspects of experimental designs, with emphasis on applied problems involved in executing a statistically sound design.

754 – Survey Sampling (3). Permission of the instructor. The different sampling techniques are discussed. Major emphasis on planning of large-scale sample surveys rather than on statistical theory.

760 – Data Collection Methods (3). Reviews alternative data collection techniques used in surveys, concentrating on the impact these techniques have on the quality of survey data. Topics covered include errors associated with nonresponse, interviewing, and data processing.

761 – Questionnaire Design (3). Examines the stages of questionnaire design including developmental interviewing, question writing, question evaluation, pretesting, questionnaire ordering, and formatting. Reviews the literature on questionnaire construction. Provides hands-on experience in developing questionnaires.

763 – Survey Computing (1). Introduces basic statistical concepts and practices emphasizing the analysis of real data. Provides training in the use of the SAS statistical analysis system and the practical problems of stratification, clustering, and weighting in survey analysis.

830 – Demography: Theory, Substance, Techniques, Part I (3). A basic introduction to the discipline of demography. Materials covered include population history, data sources, mortality and fertility trends, and differentials and techniques of analysis.

831 – Demography: Theory, Substance, Techniques, Part II (3). A continuation of SOCI 830. Materials covered include population growth and stable population theory, migration and distribution, population policy, and population estimates and projections.

832 – Migration and Population Distribution (3). Treats migration trends, patterns, and differentials and their effects on population distribution in continental and regional areas. Attention is given to theoretical and methodological problems in the study of population movement.

833 – Socioeconomic Factors in Fertility (3). Study of fertility differentials by social and economic factors, changes over time, the manner in which these factors affect fertility, and the implications thereof for fertility-control programs.

835 – Mortality: Social Demographic Perspectives (3). Prerequisite, SOCI 830. Permission of the instructor for students lacking the prerequisite. This advanced seminar covers mortality date and measurement, the inequality of death, trends in morbidity and mortality, and explanations of mortality decline. Social demographic perspectives receive primary emphasis.

905 – Survey Practicum (1). Applied workshop in sample survey design and implementation. The student works in a data collection center under the guidance of the instructor. Course focuses on real world problems in data collection and their practical, cost-effective solutions.

Data Studies Current Course Inventory: School of Information & Library Science
May, 2013
Barbara Wildemuth
School of Information and Library Science

INLS 384: Information and Computer Ethics (3). An overview of ethical reasoning, followed by discussion of issues most salient to information professionals, e.g.; intellectual property, privacy, access/censorship, effects of computerization, and ethical codes of conduct. For undergraduates only. *Offered annually. New Fall 2013.*

INLS 523: Database Systems I: Introduction to Databases (3). Prerequisite: INLS 261 for undergraduates. Design and implementation of database systems. Semantic modeling, relational database theory, including normalization, query construction, and SQL. *Offered fall and spring.*

INLS 541: Information Visualization (3). This course provides an introduction to the field of Information Visualization through readings of current literature and studying exemplars. A comprehensive review is given of the different types of information visualization techniques. The course provides students a framework for identifying the information visualization need, and determining the appropriate choice of data mappings and visualization techniques. A strong emphasis is placed on interactive electronic visualizations using freely available tools. Students will construct several visualizations as part of the class; however, no programming skills are required. There are no prerequisites. *Offered annually.*

INLS 560: Programming for Information Professionals (3). An introduction to computer programming focusing on language fundamentals and programming techniques for library and information science applications. Emphasizes problem-solving through the development of practical applications that include text processing, file handling, user interfaces, and web data access. *Offered fall and spring.*

INLS 566: Information Assurance (3). Prerequisite: INLS 261 for undergraduates. Aspects of data integrity, privacy, and security from several perspectives: legal issues, technical tools and methods, social and ethical concerns and standards. *Offered annually.*

Additional Graduate Courses

INLS 613: Text Mining (3)/ This course will allow the student to develop a general understanding of knowledge discovery and gain a specific understanding of text mining. Students will become familiar with both the theoretical and practical aspects of text mining and develop a proficiency with data modeling text. *Offered annually*.

INLS 623: Database Systems II: Intermediate Databases (3). Prerequisites: INLS 382 or INLS 582, and INLS 523. Intermediate-level design and implementation of database systems, building on topics studied in INLS 523. Additional topics include MySQL, indexing, XML, and non-text databases. *Offered fall and spring.*

INLS 624: Policy-Based Data Management (3). Students will develop policies for managing digital repositories and persistent archives. The rules will be implemented in the integrated Rule-Oriented Data System (iRODS), which organizes distributed data into shareable collections. *Offered irregularly.*

INLS 720: Metadata Architectures and Applications (3). Prerequisite: INLS 520 or 509. Examines metadata in the digital environment. Emphasizes the development and implementation of metadata schemas in distinct information communities and the standards and technological applications used to create machine understandable metadata. *Offered annually.*

INLS 722: Data Curation and Management (3). Explores data curation lifecycle activities from design of good data, through content creator management, metadata creation, ingest into a repository, repository management, access policies, and implementation, and data reuse. *Offered annually.*

INLS 723: Database Systems III: Advanced Databases (3). Prerequisite: INLS 623. Advanced study of database systems. Topics include database design, administration, current issues in development and use, optimization, indexing, transactions, and database programming. *Offered annually.*

Data Studies current course inventory: MBA
June, 2013
Sridhar Balasubramanian,
Kenan-Flagler Business School

MBA706 "Data Analytics" course description:

Pick up a recent issue of any business periodical and you are likely to find a testimonial extolling the virtues of Business Analytics, the practice of applying data and rigorous modeling to derive business insights and drive business planning. Business Analytics and its relatives Big Data, Data Mining, Business Intelligence, and Predictive Analytics represent capabilities that are increasingly sought by firms in a variety of industries

“Data Analytics” prepares students to lead analytics-driven organizations. Emphasizing fundamental data analysis concepts such as visualization, data quality, and model validation, the course will expose students to key methodological tools including linear regression, classification, clustering, and tree-based models through hands-on work with data and software. Through case discussions and a guest speaker presentation, students will critically evaluate strategic opportunities and management challenges that arise with data-driven business models. Grading will be based on several problem sets, a case writeup, and a capstone final project.

Data Studies Current Course Inventory: School of Public Health
May 2013

PUBH
450 Data Skills Online (1). This online, asynchronous class presents a series of discrete tools designed to teach skills to health professionals for using technology and data management/analysis. Online course.

754 Research Frameworks and Methods for Assessing and Improving Population Health (3). This course is designed to provide students with the fundamental research and analytic methods needed by public health leaders to assess the effectiveness, efficiency, and equity of healthcare in order to improve population health. The focus will be on research skills needed by practitioners with the objective of improving health outcomes.

741 Quantitative Methods for Health Care Professionals I (4). Permission of the instructor. Course is designed to meet the needs of health care professionals to appraise the design and analysis of medical and health care studies and who intend to pursue academic research careers. Covers basics of statistical inference, analysis of variance, multiple regression, categorical data analysis.

742 Quantitative Methods for Health Care Professionals II (2). Prerequisite, PUBH 741. Permission of the instructor. Continuation of PUBH 741. Main emphasis is on logistic regression; other topics include exploratory data analysis and survival analysis.

EPID
600 Principles of Epidemiology (3). An introductory course that considers the meaning, scope, and applications of epidemiology to public health practice and the uses of vital statistics data in the scientific appraisal of community health. One lecture and two lab hours per week.

700 SAS and Data Management (3). An introduction to statistical analysis, programming, and data management, using the SAS programming language. Two lecture hours and two lab hours per week.

710 Fundamentals of Epidemiology (4). Corequisite, BIOS 600. Permission of the instructor for nonmajors. Intensive introduction to epidemiological concepts and methods for students intending to engage in, collaborate in, or interpret the results of epidemiologic studies. An alternate to EPID 600 for satisfying the SPH core requirements. Three lecture and two seminar hours a week.

715 Theory and Quantitative Methods in Epidemiology (4). Prerequisites, EPID 705, EPID 710 or 711. Corequisite, BIOS 545. Required preparation, competence in SAS. Permission of the instructor required for nonmajors. An in-depth treatment of basic concepts and skills in epidemiologic research, including problem conceptualization, study design, research conduct, data analysis, and interpretation. Four lecture hours per week.

716 Epidemiologic Data Analysis (2). Prerequisites, EPID 705, 710 or 711. Corequisite, EPID 715. Required preparation, documented SAS proficiency. This course is a combined lecture/lab format where students get hands-on experience in the analysis and interpretation of data from cohort and case-control studies.

718 Epidemiologic Analysis of Binary Data (3). Prerequisite, EPID 715. Permission of the instructor for nonmajors. Concepts and applications, including logistic regression, binomial regression, model building strategy, additive and multiplicative interaction, and graphical exploration. Includes computer-based experience with real data. Two lecture hours and one lab hour per week.

719 Readings in Epidemiologic Methods (1). Corequisite, EPID 718 (fall); EPID 722 (spring). A discussion in journal-club format of readings in general epidemiologic methods, from problem conceptualization to application of results.

722 Epidemiologic Analysis of Time-to-Event Data (4). Prerequisite, EPID 718. Required preparation, SAS software expertise. Course covers epidemiologic analysis of time-to-event data and emphasizes weighing threats to the accuracy of inferences. Class time is spent discussing weekly readings and homeworks..

731 Systematic Review and Meta-Analysis (1). This seminar provides training in systematic review and meta-analysis. Topics include problem definition, literature search, extraction of results and study characteristics, publication bias and funnel plot analysis, analysis overall heterogeneity, and stratified and meta-regression analysis of study and population characteristics.

754 Mathematical Modeling of Infectious Diseases (3). Prerequisite, EPID 600. Introduction to basic methods for analysis and interpretation of epidemiological data on infectious diseases, and for predicting the impact of control programs such as HIV prevention programs and vaccination strategies. Two lecture hours and two lab hours per week.

BIOS

500H Introduction to Biostatistics (3). Prerequisites, MATH 231 and 232; corequisite, BIOS 511. Access to SAS, Excel required. Permission of instructor for nonmajors. Introductory course in probability, data analysis, and statistical inference designed for BSPH biostatistics students. Topics include sampling, descriptive statistics, probability, confidence intervals, tests of hypotheses, chi-square distribution, two-way tables, power, sample size, ANOVA, nonparametric tests, correlation, regression, survival analysis.

511 Introduction to Statistical Computing and Data Management (4). Required preparation, previous or concurrent course in applied statistics. Permission of instructor for nonmajors. Introduction to use of computers to process and analyze data, concepts and techniques of research data management, and use of statistical programming packages and interpretation. Focus is on use of SAS for data management and reporting.

545 Principles of Experimental Analysis (3). Permission of the instructor for nonmajors. Required preparation, basic familiarity with statistical software (preferably SAS able to do multiple linear regression) and introductory biostatistics, such as BIOS 600. Continuation of BIOS 600. Analysis of experimental and observational data, including multiple regression and analysis of variance and covariance.

550 Basic Elements of Probability and Statistical Inference I (GNET 636) (4). Required preparation, two semesters of calculus (such as MATH 231, 232). Fundamentals of probability; discrete and continuous distributions; functions of random variables; descriptive statistics; fundamentals of statistical inference, including estimation and hypothesis testing.

551 Basic Elements of Probability and Statistical Inference II (3). Prerequisite, BIOS 550. Permission of the instructor for students lacking the prerequisite. Required preparation, basic familiarity with statistical software (preferably SAS able to do multiple linear regression) or permission of the instructor. The theory and application of multiple linear regression and related analysis of variance including logistic regression and Poisson regression.

600 Principles of Statistical Inference (3). Required preparation, knowledge of basic descriptive statistics. Major topics include elementary probability theory, probability distributions, estimation, tests of hypotheses, chi-squared procedures, regression, and correlation.

660 Probability and Statistical Inference I (3). Required preparation, three semesters of calculus (such as MATH 231, 232, 233). Introduction to probability; discrete and continuous random variables; expectation theory; bivariate and multivariate distribution theory; regression and correlation; linear functions of random variables; theory of sampling; introduction to estimation and hypothesis testing.

661 Probability and Statistical Inference II (3). Prerequisite, BIOS 660. Permission of the instructor for students lacking the prerequisite. Distribution of functions of random variables; Helmert transformation theory; central limit theorem and other asymptotic theory; estimation theory; maximum likelihood methods; hypothesis testing; power; Neyman-Pearson Theorem, likelihood ratio, score, and Wald tests; noncentral distributions.

662 Intermediate Statistical Methods (4). Pre- or corequisites, BIOS 511 and 550. Principles of study design, descriptive statistics, sampling from finite and infinite populations, inferences about location and scale. Both distribution-free and parametric approaches are considered. Gaussian, binomial, and Poisson models, one-way and two-way contingency tables.

663 Intermediate Linear Models (4). Required preparation, BIOS 662. Matrix-based treatment of regression, one-way and two-way ANOVA, and ANCOVA, emphasizing the general linear model and hypothesis, as well as diagnostics and model building. Reviews matrix algebra. Includes statistical power for linear models and binary response regression methods.

664 Sample Survey Methodology (STOR 358) (4). Prerequisite, BIOS 550. Permission of the instructor for students lacking the prerequisite. Fundamental principles and methods of sampling populations, with emphasis on simple, random, stratified, and cluster sampling. Sample weights, nonsampling error, and analysis of data from complex designs are covered. Practical experience through participation in the design, execution, and analysis of a sampling project.

668 Design of Public Health Studies (3). Prerequisites, BIOS 545 and 550. Statistical concepts in basic public health study designs: cross-sectional, case-control, prospective, and experimental (including clinical trials). Validity, measurement of response, sample size determination, matching and random allocation methods.

680 Introductory Survivorship Analysis (3). Prerequisite, BIOS 661. Permission of the instructor for students lacking the prerequisite. Introduction to concepts and techniques used in the analysis of time to event data, including censoring, hazard rates, estimation of survival curves, regression techniques, applications to clinical trials.

735 Statistical Computing Basic Principles and Applications (3). Prerequisite, BIOS 661. Required preparation, familiarity with one computer system and either a computer language or computer package. Basic theory and application of computing as a tool in statistical research and practice. Topics include algorithms and data structures, linear and nonlinear systems, function approximation, numerical integration, the EM algorithm, simulation, and document preparation.

740 Specialized Methods in Health Statistics (1–21). Permission of the instructor. Statistical theory applied to special problem areas of timely importance in the life sciences and public health. Lectures, seminars, and/or laboratory work, according to the nature of the special area under study.

756 Advanced Nonparametric Methods in Biometric Research (3). Prerequisite, BIOS 661. Theory and application of nonparametric methods for various problems in statistical analysis. Includes procedures based on randomization, ranks and U-statistics. A knowledge of elementary computer programming is assumed.

760 Advanced Probability and Statistical Inference I (4). Prerequisite, BIOS 661. Permission of the instructor for students lacking the prerequisite. Measure space, sigma-field, measurable functions, integration, conditional probability, distribution functions, characteristic functions, convergence modes, SLLN, CLT, Cramer-Wold device, delta method, U-statistics, martingale central limit theorem, UMVUE, estimating function, MLE, Cramer-Rao lower bound, information bounds, LeCam's lemmas, consistency, efficiency, EM algorithm.

761 Advanced Probability and Statistical Inference II (4). Prerequisite, BIOS 760. Permission of the instructor for students lacking the prerequisite. Elementary decision theory: admissibility, minimaxity, loss functions, Bayesian approaches. Hypothesis testing: Neyman-Pearson theory, UMP and unbiased tests, invariance, confidence sets, contiguous alternatives. Elements of stochastic processes: Poisson processes, renewal theory, Markov chains, martingales, Brownian motion.

763 Generalized Linear Model Theory and Applications (4). Permission of the instructor for nonmajors. Introduction to the theory and applications of generalized linear models, quasi-likelihoods and generalized estimating equations. Topics include logistic regression, overdispersion, Poisson regression, log-linear models, conditional likelihoods, multivariate regression models, generalized mixed models, and regression diagnostics.

764 Advanced Survey Sampling Methods (3). Prerequisite, BIOS 664. Continuation of BIOS 664 for advanced students: stratification, special designs, multistage sampling, cost studies, nonsampling errors, complex survey designs, employing auxiliary information, and other miscellaneous topics.

767 Longitudinal Data Analysis (4). Prerequisite, BIOS 762. Permission of the instructor for nonmajors. Presents modern approaches to the analysis of longitudinal data. Topics include linear mixed effects models, generalized linear models for correlated data (including generalized estimating equations), computational issues and methods for fitting models, and dropout or other missing data.

772 Statistical Analysis of MRI Images (3). The course will review major statistical methods for the analysis of MRI and its applications in various studies.

773 Statistical Analysis with Missing Data (3). Prerequisites, BIOS 761 and 762. Fundamental concepts, including classifications of missing data, missing covariate and/or response data in linear models, generalized linear models, longitudinal data models, and survival models. Maximum likelihood methods, multiple imputation, fully Bayesian methods, and weighted estimating equations. Focus on biomedical sciences case studies. Software packages include WinBUGS, SAS, and R.

774 Statistical Learning and High Dimensional Data (3). Prerequisite, BIOS 661. Permission of the instructor for students lacking the prerequisite. Introductory overview of statistical learning methods and high-dimensional data analysis. Involves three major components: supervised or unsupervised learning methods, statistical learning theory, and statistical methods for high-dimensional data including variable selection and multiple testing. Real examples are used.

777 Mathematical Models in Demography (3). Permission of the instructor. A detailed presentation of natality models, including necessary mathematical methods, and applications; deterministic and stochastic models for population growth, migration.

780 Theory and Methods for Survival Analysis (3). Prerequisites, BIOS 760 and 761. Permission of the instructor for students lacking the prerequisites. Counting process-martingale theory, Kaplan-Meier estimator, weighted log-rank statistics, Cox proportional hazards model, nonproportional hazards models, multivariate failure time data.

784 Introduction to Computational Biology (3). Prerequisites, BIOS 661 and 663. Permission of the instructor for students lacking the prerequisites. Molecular biology, sequence alignment, sequence motifs identification by Monte Carlo Bayesian approaches, dynamic programming, hidden Markov models, computational algorithms, statistical software, high-throughput sequencing data and its application in computational biology.

841 Principles of Statistical Consulting (3). Prerequisites, BIOS 545. Permission of the instructor for nonmajors. An introduction to the statistical consulting process, emphasizing its nontechnical aspects.

842 Practice in Statistical Consulting (1–21). Prerequisites, BIOS 511, 545, 550, and 841. Permission of the instructor. Under supervision of a faculty member, the student interacts with research workers in the health sciences, learning to abstract the statistical aspects of substantive problems, to provide appropriate technical assistance, and to communicate statistical results.

843 Seminar in Biostatistics (1). This seminar course is intended to give students exposure to cutting edge research topics and hopefully help them in their choice of a thesis topic. It also allows the student to meet and learn from major researchers in the field.

850 Training in Statistical Teaching in the Health Sciences (1–21). Required preparation, a minimum of one year of graduate work in statistics. Principles of statistical pedagogy. Students assist with teaching elementary statistics to students in the health sciences. Students work under the supervision of the faculty, with whom they have regular discussions of methods, content, and evaluation of performance.

990 Research in Biostatistics (1–21). Individual arrangements may be made by the advanced student to spend part or all of his or her time in supervised investigation of selected problems in statistics.

HPM

470 Statistical Methods for Health Policy and Management (3). Introduction of linear model approach to analysis of data in health care settings. Topics include probability distributions, estimation tests of hypotheses, methods in multiple regression, and analysis of variance and covariance.

472 Program Evaluation (3). Concepts and methods of the program evaluation paradigm as applied in health administration.

715 Health Economics for Policy and Management (3). Prerequisite, BIOS 600. Permission of the instructor for nonmajors. Provides training in the theory of health economics and applies this theory to important issues in health policy and management.

715L Microeconomics Lab (1). Corequisite, HPM 715. Permission of the instructor for nonmajors. Applications of health economics theory to current health care policy.

742 Health Care Finance I (3). Prerequisite, HPM 740. Topics include basic financial management concepts, capital acquisition, cost of capital and capital structure, and capital allocation.

743 Health Care Finance II (3). Prerequisite, HPM 742. Topics include financial analysis and forecasting, working capital distributions to owners, mergers, capitation, and financial risk.

747 Finance and Economic Analytics (4). This course provides students in the Executive Master's Program with an opportunity to investigate topics of healthcare finance and economics in greater depth, and to extend their scope of facility with technical tools for financial and economic analysis.

760 Healthcare Quality and Information Management (3). Integrates essential methods and principles in healthcare quality and information management. Emphasis on use of information to measure and improve quality. Will include presentations, individual/group projects, exercises, and group discussion.

770 Operations Research for Health Care Systems (3). Prerequisite, BIOS 600. Permission of the instructor. Review of the systems analysis process in health care systems. Deterministic and random models, mathematical programming, queuing, simulation, forecasting, and measurement. Emphasis on model formulation and computer solution of decision models.

772 Techniques for the Economic Evaluation of Health Care (3). Prerequisite, EPID 600. This course provides an investigation of the theory, methods, and application of economic evaluation to health care. Topics include methods used to structure an economic evaluation, measure and summarize health outcomes and estimate their value to patients or to the public, and identify resources used and estimate their costs.

776 Healthcare Quality and Information Management (1.5). The HPM 776/777 and 776/778 course sequences integrate essential methods and principles in healthcare quality and information management, emphasizing use of information to measure and improve quality.

777 Health Information and Quality Applications (2). The HPM 776/777 and 776/778 course sequences integrate essential methods and principles in healthcare quality and information management, emphasizing use of information to measure and improve quality.

779 Advanced Analytics and Operations Research (4). Healthcare administrators face a range of decisions: some strategic, some financial, others operational. Through your program of study, you are developing analytical and conceptual skills that will help you to make better decisions when the time comes.

781 Seminar in Comparative Effectiveness Research (1). The course provides an overview of substantive and methodological issues in CER, including randomized controlled trials; inferences from observational studies; literature syntheses; decision sciences/decision modeling; dissemination and implementation science; cross-cutting skills (e.g., strengths and limitations of administrative and clinical databases and electronic health records for CER).

880 HPM Mathematical and Statistical Tutorial (1). Review of mathematical and statistical concepts used in HPM 881-883. Introduction to statistical programming language.

881 Linear Regression Models (3). Prerequisite, HPM 882. This course is an introduction to the analysis of categorical data using maximum likelihood. Topics covered: econometric models in which the dependent variable is not continuous, including Logit, Probit, Tobit, two-part, and duration models.

882 Advanced Methodology in Health Policy and Management (3). Prerequisites, HPM 496 and 796. This course is an introduction to linear regression models. Topics include linear algebra, least squares regression, multicollinearity, heteroscedasticity, autocorrelation, and hypothesis testing.

883 Analysis of Categorical Data (3). Prerequisites, HPM 881 and 882. Permission of the instructor for students lacking the prerequisites. Research methodology as applied to understanding problems in health care delivery. Topics include simultaneous equation models, factor analysis, limited dependent variables, and an introduction to event history analysis.

884 Health Services/Health Policy Research Methods I (3). Doctoral standing or permission of the instructor. This two-semester course provides an overview of the field of health services research and introduction to basic components of the research process, including literature synthesis, development of a research question and hypothesis, and use of conceptual and logic models to clarify research questions.

885 Health Services/Health Policy Research Methods II (3). Prerequisite, HPM 884. This continuation of HPM 884 examines basic components of the research process, including research designs, analytical issues, qualitative research methods, primary data collection, and secondary data analysis, and provides in-depth analysis of research applications that are relevant to health services and health policy researchers.

886 Advanced Applications in Research Methods (3). Prerequisites, HPM 884 and 885. This course will focus on advanced applications of research methods developed in HPM 884 and HPM 885. Examples and applications are relevant to health services and health policy researchers.

890 Special Topics in HPM (0.5-3). Course reserved for special topics in HPM for graduate-level students only.

691H Honors Research (3). Required preparation, overall grade point average of 3.2 by end of junior year in all UNC–Chapel Hill courses. Readings and seminars for undergraduates showing potential and talent for research. Students will design an independent research project, write a proposal, and complete an IRB application as partial completion of an honors thesis.

692H Independent Honors Research (3). Prerequisite, HPM 691H. Permission of the instructor. Students collect data, analyze and report findings, and make recommendations to complete an honor thesis and present findings in presentation/poster format.

ENVR

461 Environmental Systems Modeling (ENST 415, GEOL 415, MASC 415) (3). See ENST 415 for description.

470 Environmental Risk Assessment (ENST 470) (3). Required preparation, one course in probability and statistics. Use of mathematical models and computer simulation tools to estimate the human health impacts of exposure to environmental pollutants. Three lecture hours per week.

661 Scientific Computation I (MATH 661) (3). See MATH 661 for description.

662 Scientific Computation II (COMP 662, MATH 662) (3). See MATH 662 for description.

668 Methods of Applied Mathematics I (MATH 668) (3). See MATH 668 for description.

669 Methods of Applied Mathematics II (MATH 669) (3). See MATH 669 for description.

671 Environmental Physics I (3). Prerequisite, ENVR 461. A first graduate-level course in physical principles relevant to environmental systems. Topics include dimensional analysis, tensor calculus, conservation of mass and momentum. Applications are considered from natural and engineered systems and across all relevant media. Focus is on the development of mechanistic representation of environmental systems.

672 Environmental Physics II (3). Prerequisite, ENVR 671. Second part of a graduate-level sequence in physical principles relevant to environmental systems. Topics include turbulence, conservation of energy, multiscale methods, and thermodynamics. Applications are considered from natural and engineered systems and across all relevant media. Focus is on development of mechanistic representation of environmental systems.

761 Numerical ODE/PDE I (MASC 781, MATH 761) (3). See MATH 761 for description.

762 Numerical ODE/PDE II (MASC 782, MATH 762) (3). See MATH 762 for description.

763 Mathematical Modeling I (MASC 783, MATH 768) (3). See MATH 763 for description.

764 Mathematical Modeling II (MASC 784, MATH 769) (3). See MATH 764 for description.

771 Exposure Analysis (3). This course is intended for students interested in research involving exposure to environmental contaminants. The course focuses on the integration of engineering principles, with statistical tools to enhance inference. Statistical models based on the Johnson system of distributions are explored for the analysis data including exposure-biomarker relationships.

773 Modeling Atmospheric Chemistry (3). Air pollution is formed through thousands of chemical reactions. Computer models are used to simulate this complex chemistry and used to make policy. Current computational restraints force a simplified representation of atmospheric chemistry in these models, and the focus of this course is the implications of this on predictions.

783 Setting Environmental Priorities (3). This course is intended to develop a student's ability to estimate the relative merits of research and policy actions in several broad environmental areas, with attention to the associated uncertainty. Criteria to be included are both quantitative and qualitative, with an emphasis on public health, environmental, and economic metrics.

850 Systems Analysis in Environmental Planning (3). Required preparation, calculus. Applications of systems analysis techniques to the management of environmental quality.

890 Problems in Environmental Sciences and Engineering (1–21). Permission of the department. For students outside the department who wish to undertake individual study of a specific problem in environmental sciences and engineering. The subject and requirements of the project are arranged with the faculty in each individual instance. One or more hours per week.

601 Epidemiology for Environmental Scientists (3). An introduction to relevant epidemiologic concepts that inform environmental science research. Learning objectives include discussing basic epidemiologic concepts and measures of disease occurrence in populations, explaining epidemiological study designs for studying associations between risk factors or exposures in populations, evaluating epidemiologic evidence, and comprehending basic ethical principles.

755 Analysis of Water Resource Systems (3). Permission of the instructor for nonmajors. Use of mathematical models to design and evaluate regional water supply and treatment systems. Engineering and economic methods are incorporated into quantitative analyses of regional scenarios. Social and political aspects also discussed. Three lecture hours per week.

640 Environmental Exposure Assessment (3). Permission of the instructor for nonmajors. The course material introduces the general concepts of assessing environmental exposures to chemicals in human populations. This includes the design of ecologic and personal monitoring studies, the techniques and equipment used for sampling and analysis, and interpretation of data.

666 Numerical Methods (3). Prerequisites, COMP 116 and MATH 383. Numerical methods for solving problems arising in sciences and engineering. Solution of linear equations using direct and iterative approaches, solution of nonlinear systems of algebraic equations, solution of ordinary differential equations including single and multistep methods, and methods for stiff systems of ODEs and collocation methods for linear and nonlinear PDEs.

570 Methods of Environmental Decision Analysis (3). Required preparation, one course in probability and statistics. Use of quantitative tools for balancing conflicting priorities (such as costs versus human health protection) and evaluating uncertainties when making environmental decisions.

MHCH

712 Program Assessment in Maternal and Child Health (3). Permission of the instructor for nonmajors. Offers an opportunity for students to explore in greater depth a selected MCH practice topic. Students will learn how to provide consultation about a selected program activity.

713 Research Methods in Maternal and Child Health (3). Permission of the instructor for nonmajors. The art and science of MCH research, with an emphasis on applied survey research. Student groups will design and carry out a small survey, and present their findings in a poster presentation. Focuses on assessment of MCH population characteristics, secondary data analysis, and the evaluation of MCH programs. A practicum-based course. Three lecture hours per week.

713L Research and Evaluation Methods in Maternal and Child Health Lab (1). Corequisite, MHCH 713. Permission of the instructor for nonmajors. The MHCH 713 lab, which is a companion course to MHCH 713, introduces students to statistical analysis using SPSS-PC and microcomputers. Two lab hours per week.

862 Maternal and Child Health Program Evaluation (3). Knowledge of Stata or similar statistical package required. Analytic skills seminar on the theory and practice of program impact evaluation. Topics: what is impact evaluation; key issues to consider when evaluating program impact; selectivity and other problems when evaluating program impact; research designs and estimation strategies; interpretation of results.

NUTR

809 Applied Qualitative Research Methods (2). Introduces students to qualitative research methods with an emphasis on their use in nutrition-related programmatic research. Uses a combination of didactic, interactive, and applied techniques to teach qualitative research knowledge and skills. Introductory course; subsequent courses in qualitative methods, particularly for data analysis, is strongly recommended.

815 Diet and Cancer (EPID 815) (3). Prerequisites, BIOS 600, EPID 600 or 710, 771, and NUTR 813. Examines and critically evaluates epidemiologic research on relationships of diet-related exposures with cancer etiology, prevention, and survivorship. Emphasis on skills for conducting, analyzing, and interpreting diet and cancer epidemiologic studies.

818 Analytical Methods in Nutritional Epidemiology (EPID 818) (3). Prerequisites, BIOS 545, EPID 600 or 710, and NUTR 813. Skills and techniques to study how dietary exposures, physical activity, and anthropometric status relate to disease outcomes. Focus is hands on data analysis using STATA, and interpretation of results from statistical analysis.

845 Nutritional Metabolism (3). Prerequisite, NUTR 600. A problem-based approach to examine current topics in biochemistry relevant to nutrition and metabolism. Students interpret data and design experiments related to recent advances in nutritional biochemistry.

885 Doctoral Seminar (1). This course is designed for doctoral and master of science students only. Critical review of current literature in nutritional biochemistry, intervention and policy, and population-based nutrition science. Focuses on the development of skills in reviewing and criticizing articles.

HBEH

601 Principles of Statistical Inference for Health Behavior (3). Required preparation, knowledge of basic descriptive statistics. Majors only. Major topics include elementary probability theory, probability distributions, estimation, tests of hypotheses, paired and independent samples t-tests, ANOVA, linear and logistic regression, correlation and chi-squared procedures. SAS, a statistical software package, is used in the course.

743 Program Intervention, Implementation, and Monitoring II (1–4). Prerequisite, HBEH 742. Application of methods to analyze and interpret data regarding the effectiveness of health education interventions. Students work under faculty advisors to assess the effectiveness of interventions implementation in HBEH742.

744 Research Practicum I (2). Students must complete a mentored research practicum. The mentor and student will develop a contract to achieve the research. The practicum requires a total of two hundred hours of work starting in the second year of the program.

745 Research Practicum II (2). Prerequisite, HBEH 744. Students must complete a publishable manuscript based on the Research Practicum I course.

750 Applied Research Methods (3). Permission of the instructor for nonmajors. Research methods of relevance to planned change in health-related behavior and program planning. Research designs include quantitative and qualitative methods and focus on application to public health practice. Four lecture hours per week.

753 Qualitative Research Methods (NUTR 753) (3). Prerequisite, HBEH 750. Approaches to designing qualitative research studies for the development and evaluation of public health programs. Emphasis is on the practice of collecting and analyzing data from individual interviews, focus group discussions, and observations.

754 Advanced Qualitative Research Methods in Health Behavior and Health Research (3). Prerequisite, HBEH 753. This course provides advanced graduate students in public health and related fields the opportunity to explore different analytic approaches and techniques and develop analysis and writing skills. Students will apply methods they learn to analyze, interpret and write-up the results of their own qualitative research.

760 Advanced Research Methods I (3). Permission of the instructor for nonmajors. Doctoral seminar on fundamentals of research in health behavior, including conceptualization of research questions and hypotheses, measurement, sampling, and observational research designs.

761 Advanced Research Methods II (3). Prerequisite, HBEH 760. Permission of the instructor for nonmajors. Doctoral seminar on sampling and selected topics in statistical analysis; continuation of HBEH.

850 Research Manuscript Development (3). Prerequisite, HBEH 751 or 860. This seminar is designed to help advanced students refine conceptual and writing skills essential to the production of a manuscript based on already collected qualitative and quantitative data. Three hours per week.

851 Causal Modeling and Structural Equations (3). Prerequisite, BIOS 545. Permission of the instructor. This seminar is designed to refine a wide range of research skills in health behavior by using data collected by others. Three seminar hours per week.

852 Scale Development Methods (3). Prerequisite, HBEH 750. Permission of the instructor. Covers theory and application of scale development techniques for measuring latent constructs in health research; classical measurement theory and factor analytic methods are emphasized. Three seminar hours per week.

PROPOSED TEACHING AND OTHER OPPORTUNITIES IN BIG DATA FOR KENAN-FLAGLER BUSINESS SCHOOL

This document was prepared by Noel Greis and shared with Senior Associate Dean Jennifer Conrad in the context of discussions in June 2013 about how to integrate "big data" into the fabric of the Business School.

I. TEACHING

According to a recent McKinsey report, by 2018 "there will be a need for 1.5 million more data-savvy managers to take full advantage of big data in the U.S." Kenan-Flagler MBAs and undergraduates could benefit from electives that look at how big data and its tools are transforming business.

Some high-level ideas for an elective (s) that we could offer are sketched below. These electives would focus on the "business" side of big data rather than the "science" side and would be designed for students without deep technical skills.

Big Data & Business (or Digital Business). This course would look at how big data is having an impact on all aspects of an organization. Each week would be devoted to a specific business process/functional area such as marketing (customer purchasing patterns), operations ("smart" supply chains), finance (evaluating financial/economic risks) or particular industries such as health care (cost-saving strategies for healthcare allocation). The goal is to use case studies to explore how organizations are using big data effectively.

Exploiting Web Intelligence. The elective would take an in-depth look at how information on the web can be mined for business intelligence that can be used specifically to create new business ideas. The course is designed for entrepreneurial students who are interested in exploiting big data sources arising from social media, mobile devices and sensors to develop applications and other new products and services.

Managing in a Big Data World. This course would look at how big data is transforming organizational management and ideas about leadership. It would explore how traditional decision-making is affected by the use of data to drive actionable insights that lead to major corporate decisions. It could also address how an organization positions "big data" within the organization's structure and the constraints that might be needed to manage in the big data world.

Big Data Analytics for Health Care.[1] Health care organizations are generating petabytes of data from patient care data to health care payer data. As health care moves toward "meaningful use" and "evidence-based medicine," and as pressures for lowering the cost of health care increase, new big data analytics are playing an increasingly large role. This course would look at specific applications of big data tools and predictive analytics that span the health care delivery system.

II. INTELLIGENT SYSTEMS LAB

The Intelligent Systems Lab could serve as a "hands-on" resource for students to play with big data analytics. While small, the lab could support student projects, dissertations, senior theses, etc. It could also serve as a "lab" for exercises that are part of the elective courses. The lab consists of a suite of three rooms/offices on the 4th floor of the Kenan Center. This space contains the servers, monitors, and other equipment that support our corporate and other projects. Monica Nogueira, Aurelio DiScala and Holton Thompson have offices in this space.

III. KENAN-FLAGLER "*DATAJAM*": A NATIONAL CASE COMPETITION

Modeled after the North Carolina ***DATAJAM***, this student competition would bring student teams from across the country to Kenan-Flagler for an event that combines national/corporate speakers in the area of big data with a student competition. Competing teams would formulate big data solutions that catalyze new business opportunities or projects with a positive community/economic impact. The ***DATAJAM*** could be a high-profile national event that brings attention to Kenan-Flagler and UNC. We could look to some of the local "big data" companies for financial support.

IV. KENAN_FLAGLER INTERNATIONAL SYMPOSIUM ON BIG DATA AND BUSINESS

An international/national symposium on "big data" could provide visibility for Kenan-Flagler and UNC. A number of different formats are possible. But, generally, such a symposium would attract companies interested in hearing from other companies about their experiences implementing "big data" within their organizations, as well as academics and technology providers who are developing new analytical tools. This symposium might be designed in parallel with a similar symposium on "Internet-of-Things" that is being planned for the UNC-Tsinghua Center for Logistics and Enterprise Development for late 2014.

[1] Noel Greis and Monica Nogueira are planning to teach this course as an MBA elective in the 2014-2015 academic year.

V. EXECUTIVE EDUCATION CERTIFICATE/PROGRAM IN BIG DATA

Our expertise working with companies in the big data arena would provide a foundation for helping to launch such a program. [Incidentally, we just started to collaborate with a health care executive who is part of the MBA@UNC program on a project to apply machine learning to the problem of helping hospitals manage denied claims.]

Curricular Opportunities in "Big Data"

A Dean's Fellows Project by:

Lara Koch
Jon Holbrook
Chris Huchenski
Parker Mcallister

March 28, 2014

UNC KENAN-FLAGLER BUSINESS SCHOOL

The Team

JON HOLBROOK

Davidson College
Pre-MBA Experience:
Teach for America, New Orleans
Senior Marketing Specialist, LivingSocial

CHRIS HUCHENSKI

Bucknell University
Pre-MBA Experience:
Retail & Financial Planning, Lord & Taylor

LARA KOCH

Duke University
Pre-MBA Experience:
Equity Trading, Lehman Bros. and Barclays Capital
Founder, MacroCatalysts

PARKER MCALLISTER

Elon University
Pre-MBA Experience:
Entrepreneurship & Business Development, Shanghai, China

Agenda

- Introduction
- Recruiters
- Competitors
- Students
- Next Steps

What is "Big Data"?

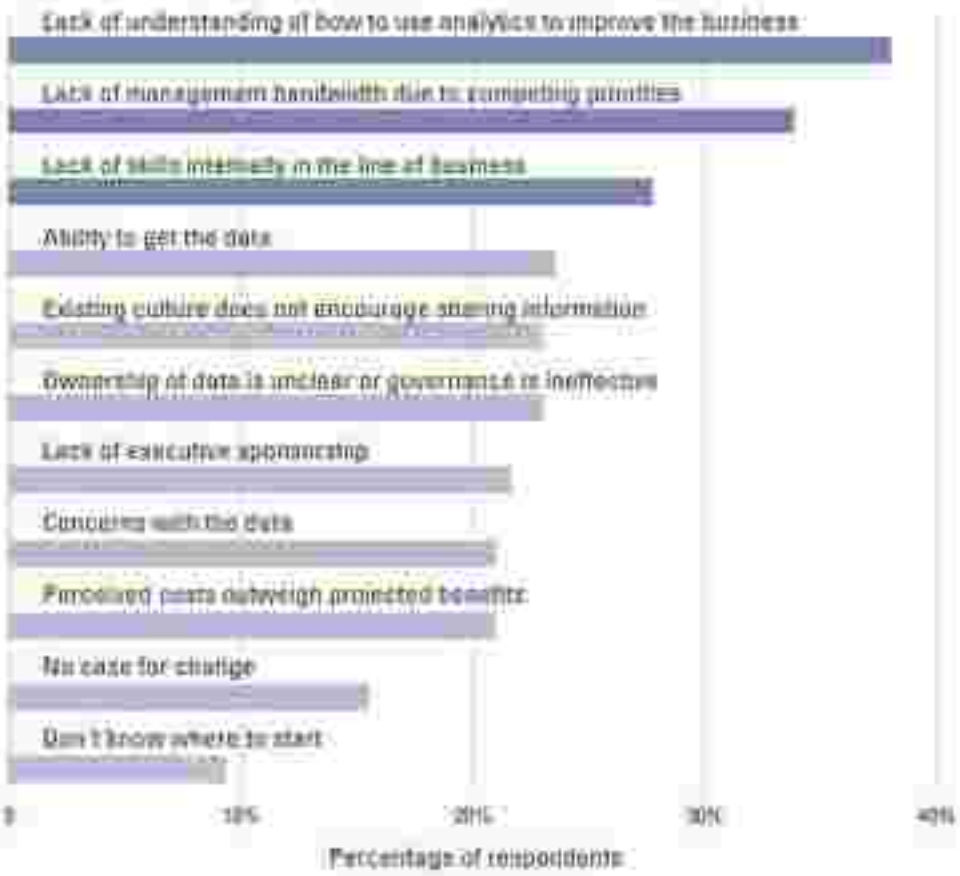
Rapid expansion in the volume, variety, and velocity of data has fueled the recent trend in Data Analytics

Volume	Data volume is increasing at a rate of 2.5 exabytes (or 2.5 **billion** GB) **daily.** ***Example:*** *the average company in 15 of 17 U.S. sectors now has more data stored than the Library of Congress.*
Variety	Content comes from more diverse sources and formats. ***Example:*** *data can be sourced from anything from a smartphone, a Tweet, a GPS device, a Google search, a parking meter, etc.*
Velocity	Faster analytics unlock value and reduce risk. ***Example:*** *organizations can monitor hundreds of real-time video feeds concurrently to more quickly identify a potential security threat.*

Sources: Boston Consulting Group and McKinsey & Co.

Why is Big Data relevant to KFBS?

Most Common Barriers to Widespread Adoption of "Big Data" within an Organization*



1. Businesses need **Data Leaders,** not **Data Scientists**.
2. Universities have started to respond… yet Kenan-Flagler still has **an opportunity to differentiate** itself as a first-mover amongst MBA programs.
3. The **demand from current and prospective students** for more Data Analytics coursework is high.

Source: MIT Sloan School of Management

Our Recommendation

We recommend adding an enrichment concentration in **Data Analytics and Decision-Making** to the current MBA program.

A concentration offers the most upside* for **Kenan-Flagler** and its **MBAs**, as well as for the broader **business community.**

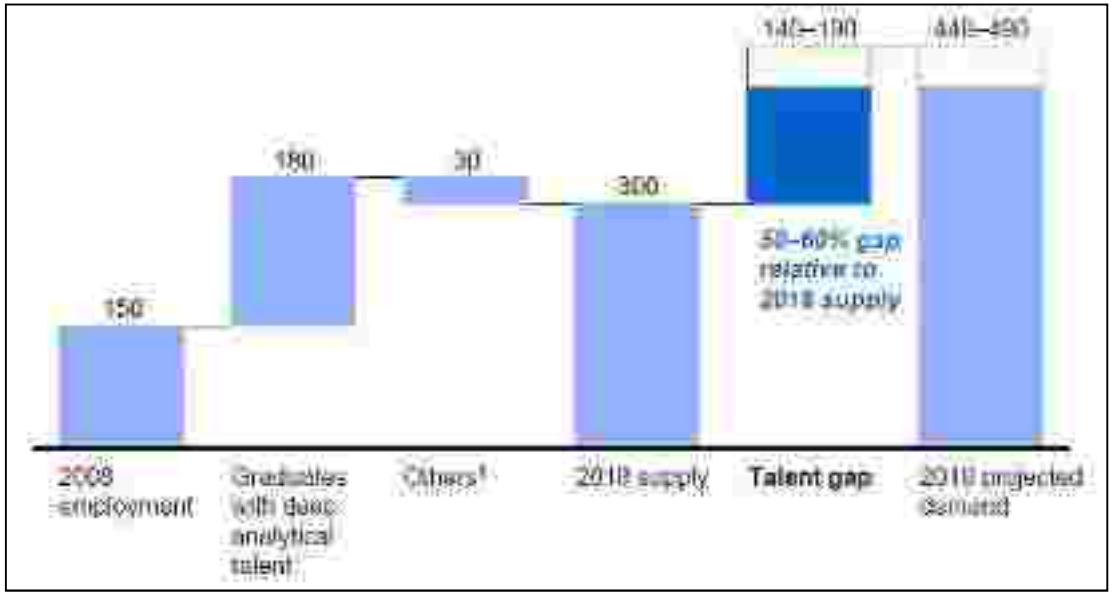
** For details on other alternatives considered, please see Appendix.*

Organizations that have adopted data analytics tend to outperform their peers

Top-performing organizations are more likely to use analytics to inform daily decision-making

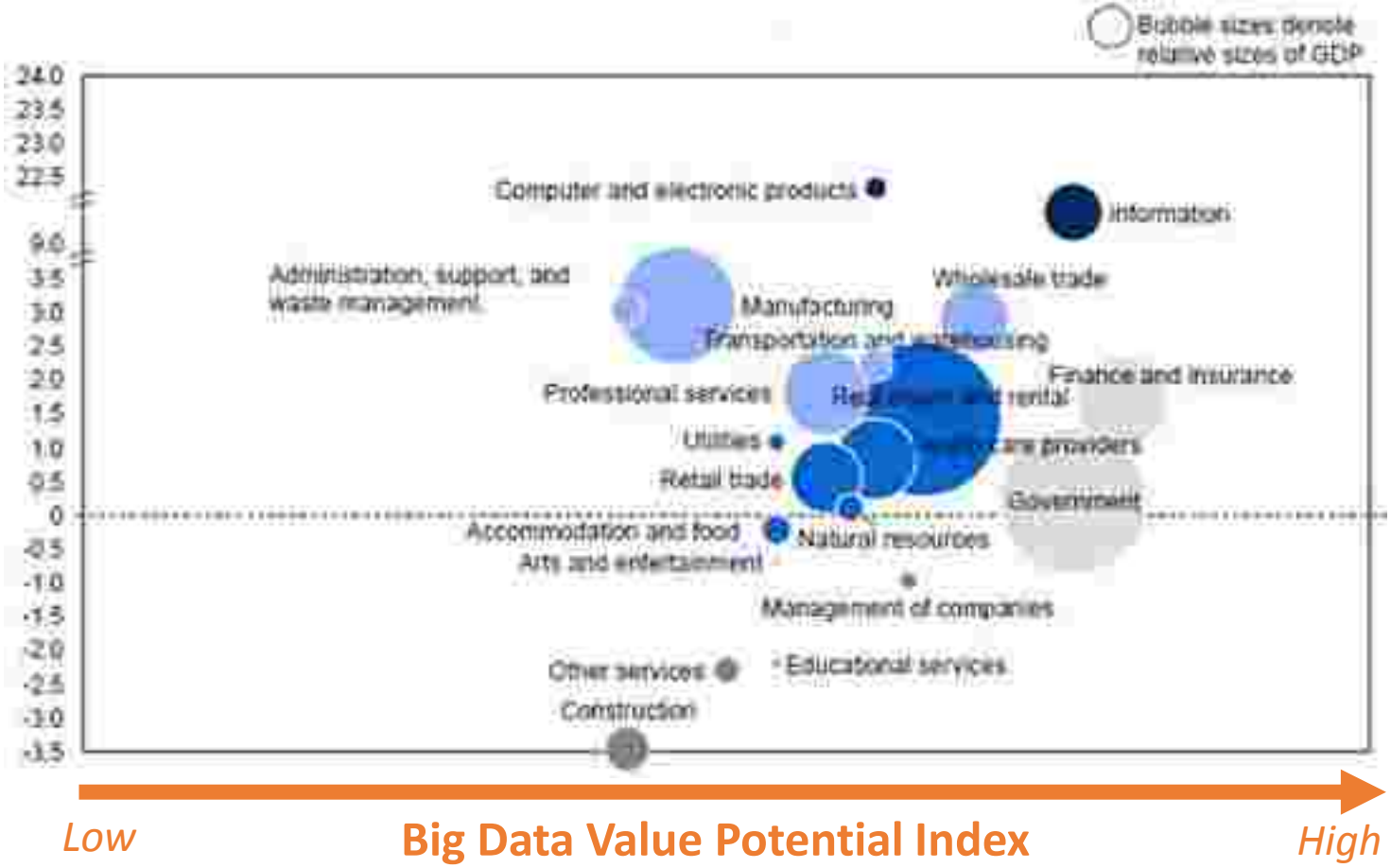
Talent gap expected to widen as Big Data becomes more accessible

By 2018, U.S. will need an additional 1.5M managers and analysts "who can ask the right questions and consume the data effectively"



Source: McKinsey & Co.

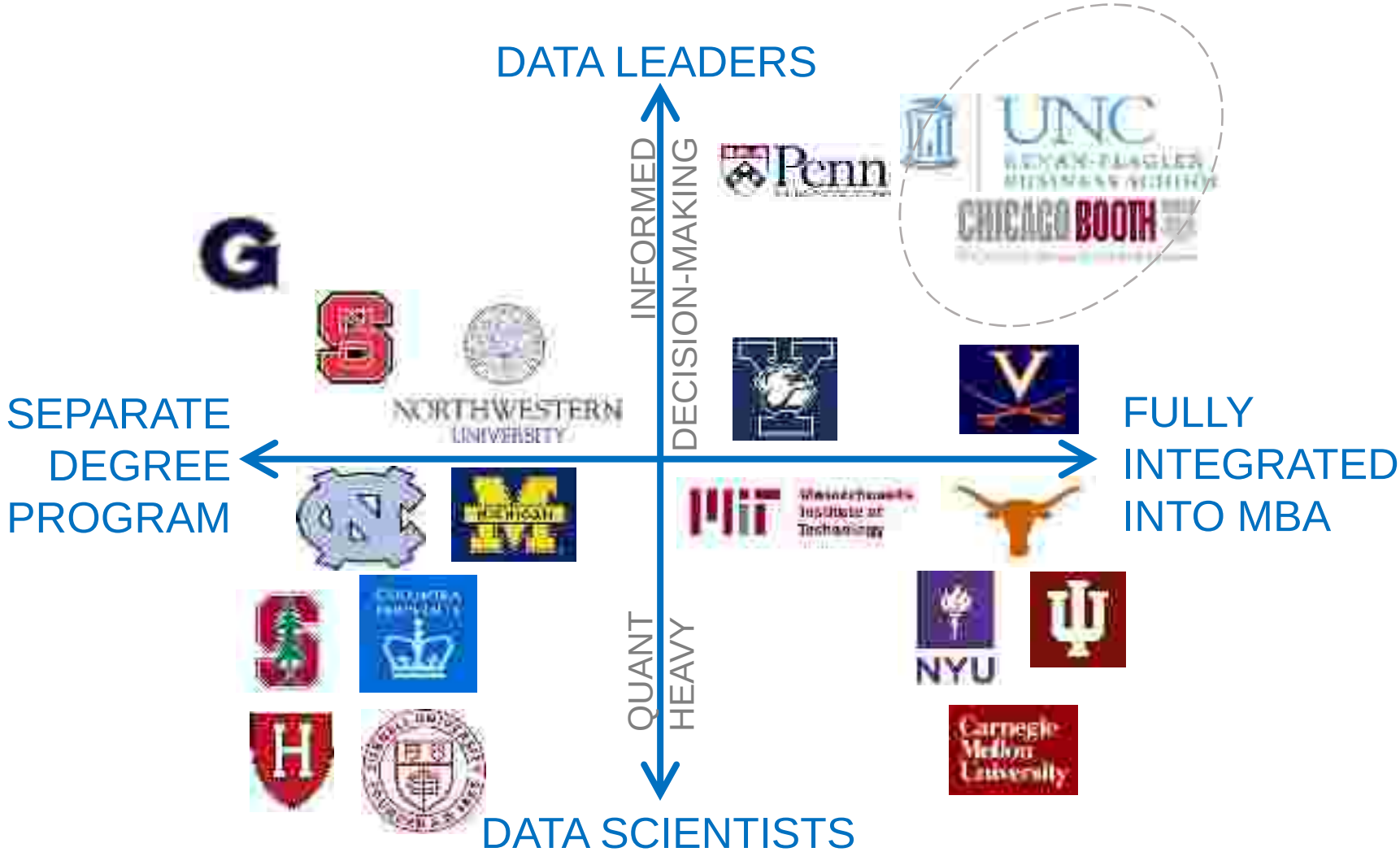
Enrichment would strengthen, as well as diversify, on-campus recruiting



Low **Big Data Value Potential Index** *High*

Source: McKinsey & Co.

UNC can fill unique space among peers



Chicago Booth's Data Analytics concentration offers the best template

SCHOOL	PROGRAM MARKETING	COURSE OFFERINGS	PROGRAM RANKING	COST	TOTAL
CHICAGO BOOTH	●	●	●	◔	●
	●	◕	◔	●	◕
	◕	●	◔	◕	◕
NYU	◑	◑	◕	◑	◑
	◔	◔	◑	◑	◔

● More favorable to incoming students ◔ Less favorable to incoming students

Data Analytics elective yields waitlist of more than 100 students

Adam Mersereau's new elective, "Data Analytics: Tools and Opportunities," had as many as 110 students on the waitlist prior to Mod IV

A week into Mod IV, that waitlist still stands 50-students strong…

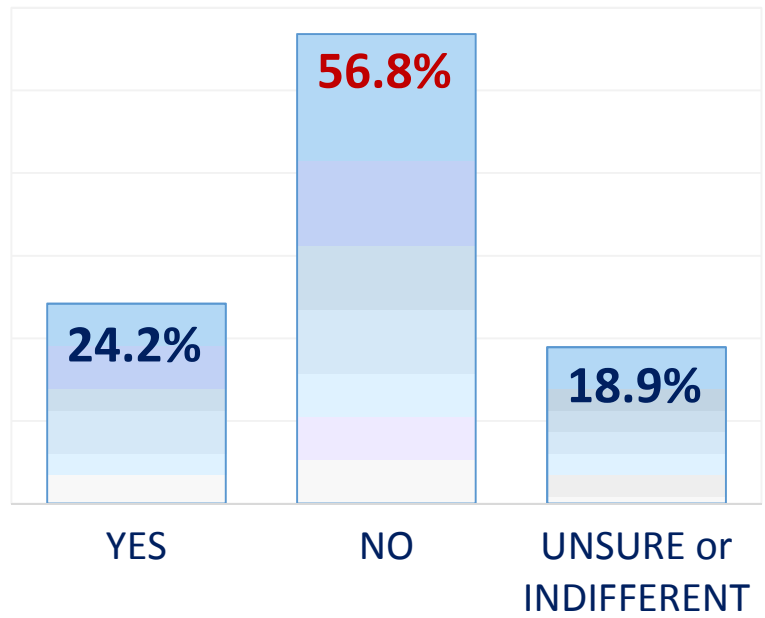
Number	Title	Instructor	Type	Meetings	Module	Enrollment Status
MBA706-001	Data Analytics: Tools and Opportunities	Adam Mersereau	Elective	TR 2:00PM- 3:20PM MB 2650	Module IV	0 / 45 Wtlst 50

CONTEXT: *at this point in the Mod, only one other course has a waitlist……* **of 7**

Current MBAs want more opportunities to learn about Data Analytics

Results from our survey* of over 220 current MBAs suggest that the demand for Data Analytics coursework is both significant and widespread.

Do you think Kenan-Flagler currently offers enough opportunities to learn about Data Analytics?



*Detailed survey results can be found in the Appendix.

Admitted students have expressed interest

Hi Lara,

My name is [Rambo] and we ***met at Fellows weekend*** *a few weeks ago.* ***I wanted to follow up with you about your proposal to add more quantitative/analytic courses to the curriculum and see what kind of reception you got from the Dean and faculty at Kenan-Flagler.***

I also wanted to get your perspective on the Dean's Fellow Program and how that has contributed to your experience at KF. I am applying for the program now and really like how it offers me an opportunity to leave a lasting impact on the community.

I look forward to hearing from you.

Thanks,

Rambo

An enrichment would be ideal complement to existing career concentrations

CAREER CONCENTRATIONS
Corporate Finance
Capital Markets and Investments
Management Consulting
Marketing
Operations Management
Real Estate

ENRICHMENT CONCENTRATIONS
Data Analytics and Decision-Making
Entrepreneurship
Healthcare
Sustainable Enterprise

Curriculum should resemble format of existing concentrations and offer blend of current and new courses

DATA ANALYTICS AND DECISION-MAKING (7.5 credits)
Two (2) Required Courses (3 credits)
Data Analytics: Tools and Opportunities
World Class Decision-Making Skills Capstone [*New Course*]
A Minimum of Three (3) Elective Courses (4.5 credits)
Marketing Analytics
Pricing
Operations Management Models
Big Data Analysis and Research Applications [*New Course*]
Introduction to Data Mining [*New Course*]

KFBS should leverage existing resources for enrichment program

Identify Potential Faculty Interested in Developing & Supporting Analytics Curriculum

Dr. B, Albert Segars, Adam Mersereau, Arvind Maholtra, Tarun Kushwaha, Wendell Gilland, Alan Neebe, Matt Pearsall, etc.

Partner with SAS, IBM, & Quintiles

1. Guest Lecturers, adjunct professor(s)
2. Use live cases from companies for class
3. Build stronger relations for recruiting

Concluding Remarks

• Differentiate KFBS from other MBA programs • Complement current strengths (i.e. Leadership Initiative, Experiential Learning) • Increase brand awareness	**KENAN-FLAGLER**
• Meet demand of current MBAs • Increase conversion of top talent from applicant pool • Equip MBAs for evolving business landscape • Enable promotion of analytical skills via résumé	**STUDENTS**
• Meet growing demand for Data Leaders vs. Data Scientists • Expand network of on-campus recruiters	**RECRUITERS**

Multiple stakeholders would benefit from a Data Analytics concentration.

Appendix

Original Project Assignment

- Analyze increasing relevance of data analytics across industries
- Examine how competitor programs have responded to “big data” trend
- Compare those responses to current Kenan-Flagler curriculum

Recommend how Kenan-Flagler should integrate Data Analytics into curriculum going forward

Alternatives Considered

Alternative	Pros	Cons
An enrichment concentration in Big Data Analytics	Provides ample time for a well-rounded understanding of Big Data, most bang for buck with recruiters, most popular amongst students	Hardest to implement – may require new courses and hiring new professors to teach those courses
One-day conference on Big Data	Students can use it as an opportunity to "dip their toes" into Big Data	Not enough time to master a difficult subject, ultimately doesn't pay dividends
Interactive seminar (i.e. Training the Street) focused on Big Data	Relatively easy to implement, open to all students with even slight interest	Impossible to learn enough in a one-day seminar to create value in recruiting
A speaker series focused on Big Data	Relatively easy to implement, open to all students with even slight interest	Material is difficult to learn without doing; could be hard to find speakers
An MBASA Club for Big Data enthusiasts	Flexible, student-run, and has the ability to evolve every year	Big Data is very technical, and requires more hands-on classroom time
Cross-disciplinary course partnering with broader UNC community	Easy to implement	Little benefit realized during recruiting cycle
A Big Data case competition	Students gain the experience of solving real-world problems	Possible that only winners differentiate themselves

Current Student Survey Results

Survey of current 1st and 2nd year MBAs (February 2014)

Question 1. On a scale from 1-5, how familiar are you with the topic of "Big Data"?

Min Value	Max Value	Average Value	Standard Deviation
1.00	5.00	2.85	0.88

Question 2. On a scale from 1-5, in your prior job(s), how often did you analyze large quantities of data and/or rely on large quantities of data to inform your decision-making?

Min Value	Max Value	Average Value	Standard Deviation
1.00	5.00	3.16	1.06

Question 3. Do you think Kenan-Flagler currently offers ample opportunities to learn about Data Analytics?

Yes	24%
No	57%
Don't Know/Don't Care	19%
Total	100%

Question 4. On a scale from 1-5, please indicate how interested you'd be in each of the following items if they were offered by the program.

#	Answer	Min Value	Max Value	Average Value	Standard Deviation
1	A Speaker Series on Data Analytics in the Workplace (either as part of a one-day conference dedicated to Big Data, or through a series of lunch sessions, etc.)	**1.00**	**5.00**	**3.18**	**1.08**
2	An interactive teach-in on the practical applications of Data Analytics (i.e. similar in structure to Training the Street classes).	**1.00**	**5.00**	**3.74**	**1.09**
3	An academic enrichment in Data Analytics	**1.00**	**5.00**	**3.10**	**1.25**
4	An MBASA club dedicated to Data Analytics.	**1.00**	**5.00**	**2.16**	**1.06**
5	A cross-disciplinary course that partners with other programs at UNC.	**1.00**	**5.00**	**3.06**	**1.12**
6	Case competition dedicated to Big Data.	**1.00**	**5.00**	**2.73**	**1.20**

Question 5. Which year are you currently in of the MBA program?

Answer	%
1st Year	61%
2nd Year	39%

Question 6. In which academic areas are you specializing? (please check all that apply)

Answer	%
General Management	26%
Corporate Finance	35%
Capital Markets and Investments	9%
Management Consulting	23%
Marketing	30%
Operations Management	16%
Real Estate	12%
Entrepreneurship	17%
Healthcare	11%
Sustainable Enterprise	14%
Energy	8%
Family Business	3%
Environment	2%
Law	2%
Public Health	2%

Screenshots of Competitor Programs

Demonstrative of the various forms of (a) curricular responses to “Big Data” and (b) marketing strategies to promote those programs

Competitor Profile

University of Chicago

WHO: Booth School of Business

WHAT: Business Analytics concentration within broader Booth 2-year MBA program

- **Class Size:** 581
- **Ranked #9** globally by the Financial Times
- **In-state tuition:** N/A
- **Out of state tuition:** $117,520 for two years

http://www.chicagobooth.edu/programs/full-time/academics/curriculum/analytic-management#tabContent2

Competitor Profile

University of Texas

WHO: McCombs School of Business

WHAT: Business Analytics concentration within broader McCombs two year MBA program

- **Class Size:** 275
- **Ranked #39** globally by the Financial Times
- **In-state tuition:** $68,000 for two years
- **Out of state tuition:** $96,000 for two years

Curriculum Requirements

Required courses:

- MKT 282 – Analysis of Markets (fall of first year)
- STA 287 – Business Analytics and Decision Modeling (spring of first year)

Any four courses from the following:

- MIS 382N 13 - Data Mining for Business Intelligence (spring of first or second year, preferably first year)
- MIS 382N 12 - Social Media Analytics (spring of second year)
- MIS 382N - Decision Support Modeling (spring of second year - either STA 287 or MIS 382N)
- STA 380 - Business Analytics: Advanced Applied Regression
- OM 382 - Supply Chain Management (spring of second year)
- OM 383 - Pricing and Revenue Management (fall of second year - currently not offered)
- MKT 382 - Marketing Information and Analysis (fall of second year)
- MKT 382 - Marketing Decision Making in the Information Age (spring of first or second year)
- MKT 382.13 - Marketing Metrics (spring of second year)

Additional courses may be approved on a course-by-course basis.

Total credit hours: 10

Faculty Contact: Prabhudev Konana

http://www.mccombs.utexas.edu/MBA/Full-Time/Academics/~/link.aspx?_id=E049C5B7E4D547269EE3BD6B011256BF&_z=z

Competitor Profile

Indiana University

WHO: Kelley School of Business

WHAT: Business Analytics major within broader Kelley two year MBA program

- **Class Size:** 204
- **Ranked #47** globally by the Financial Times
- **In-state tuition:** $54,148 for two years
- **Out of state tuition:** $93,100 for two years

http://kelley.iu.edu/mba/academics/MajorsMinors/page38907.html

Competitor Profile

New York University

WHO: Stern School of Business

WHAT: Business Analytics specialization within broader Stern two year MBA program

- **Class Size:** 392
- **Ranked #17** globally by the Financial Times
- **In-state tuition:** N/A
- **Out of state tuition:** $122,548 for two years

NYU Stern | Full-time MBA |...

Business Analytics

The specialization in Business Analytics teaches the use of data and models to support decision making in business. Students learn how to model such relationships as the impact of advertising on sales, how historical data predict stock returns, and how changes in task characteristics can influence time to completion.

Sample Courses:

- Data Mining for Business Intelligence
- Data Visualization
- Decision Models
- Econometrics
- Forecasting Time Series Data
- Introduction to Stochastic Processes
- Managing for Quality
- Risk Management Systems
- Trading Strategies and Systems
- Regression and Multivariate Data Analysis

Especially Applicable to Careers in Consulting, Finance and Marketing such as:

- Brand Management
- Internal Consulting
- Investment Banking
- Management Consulting
- Marketing
- Operations
- Public Finance
- Real Estate Finance
- Risk Management
- Strategic Planning

http://www.stern.nyu.edu/programs-admissions/full-time-mba/academics/specializations/business-analytics/

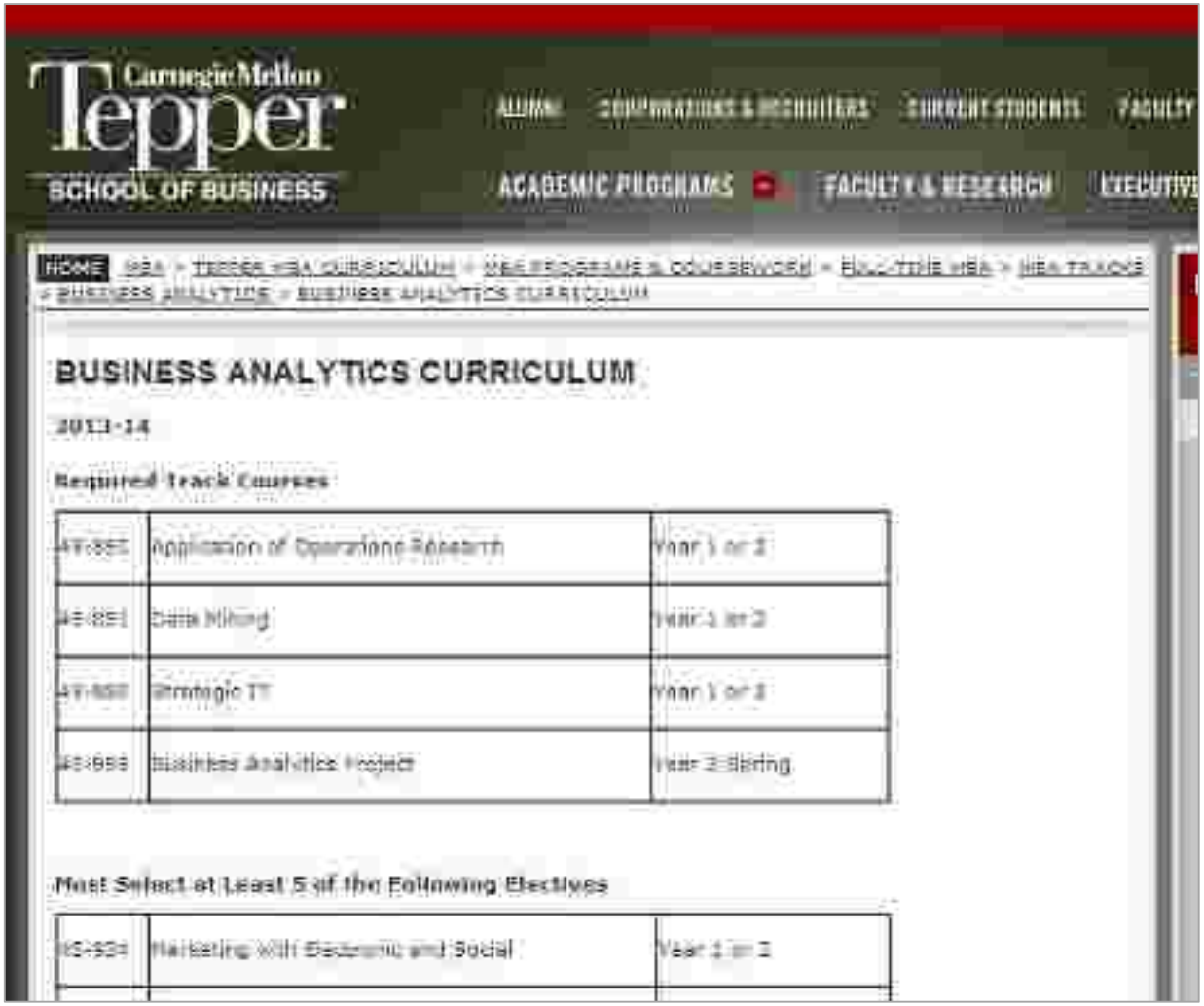
Competitor Profile

Carnie Mellon University

WHO: Tepper School of Business

WHAT: Business Analytics track within broader Tepper two year MBA program

- **Class Size:** 201
- **Ranked #34** globally by the Financial Times
- **In-state tuition:** N/A
- **Out of state tuition:** $116,600 for two years



http://www.tepper.cmu.edu/mba/mba-curriculum/mba-programs-coursework/fulltime-mba/mba-tracks/business-analytics/index.aspx

Competitor Profile

University of Virginia

WHO: Darden School of Business

WHAT: Market Analytics theme track concentration within broader Darden two year MBA program

- **Class Size:** 316
- **Ranked #27** globally by the Financial Times
- **In-state tuition:** $101,800 for two years
- **Out of state tuition:** $111,800 for two years

Theme Track Concentrations

Corporate Innovation

Students fulfilling this concentration graduate with the mindset and skill set to succeed at identifying and implementing innovation in a corporate setting. Faculty Contact: [illegible]

Business Development and Growth

Sustainable success requires organizations to innovate and grow to maintain or increase profitability. Students focusing on this concentration will learn the fundamentals of managing innovation, organic growth, and acquisition and integration of mergers and acquisitions. Faculty Contacts: Mike Lenox and Ed Hess.

Market Analytics

The objectives of the market analytics concentration are to equip students with strong analytic skills, expose them to large data sets where they can engage in mining as it relates to customer relationship management, brand marketing, and other processes for building closer relationships with one's customer base. The concentration is positioned for those students who wish to follow a B2C career with more of a focus on the research, who are looking at careers in consulting, business development, or other areas where analytic skills are highly valued. Faculty Contact: [illegible]

Sustainability

The objectives of this concentration are to provide students with knowledge about the global impacts on natural systems from rising energy and material consumption patterns; to inform students of institutions and policy instruments that influence business operations; to expose them to innovative processes and tools available; and to provide guidance on how innovative sustainability practices can be designed to increase revenues and lower costs. View a video about Darden's sustainability curriculum. Faculty Contact: Ulrich Lenz

Internal Profile

University of North Carolina

WHO: Department of Statistics and Operations Research

WHAT: a stand-alone masters degree in Interdisciplinary Operations Research and Statistics; Business Analytics is one of several "tracks" offered through this program

NOT PART OF THE MBA PROGRAM, but could leverage resources from this program's curriculum to supplement KFBS Data Analytics Courses if need be…

Track II: Business Analytics

Brief Description:

Decision making by businesses and organizations is increasingly driven by data. Examples are found in revenue management, marketing, fraud detection, supply chain management, health care management, disease control, traffic congestion control etc. Organizations collect large amounts of data and use it to make intelligent decisions. This requires strong mathematical modeling skills as well as solid background in statistical data analysis and optimization. UNC STOR department offers a unique combination of courses that a student can take to build expertise in these areas. In addition, students can augment these courses by selecting appropriate courses from the Business School, School of Public Health, the Economics department, and the Computer Science department. This allows our students to specialize in a particular area of interest in addition to getting strong methodological training.

The students enrolled in this track are expected to take the following three courses in the first semester:
1. STOR 656: Mathematical Statistics
2. STOR 613: Deterministic Models in Operations Research I
3. STOR 641: Stochastic Models in Operations Research I

In addition, the students must take seven electives and one presentation course (STOR 790). At least four of the seven electives courses must be STOR courses at least two of which must be at level 600 and above. The student may select further electives from other departments such as Business School at UNC, Business School at Duke, Computer Science, Economics, etc. However, all course selections must be approved by the graduate advisor.

Curriculum details continue on following slide …

Internal Profile: University of North Carolina *(continued)*

MS in Business Analytics through the Department of Statistics and Operations Research

REQUIRED CURRICULUM

The students enrolled in this track are expected to take the following three courses in the first semester:

1. STOR 555: Mathematical Statistics
2. STOR 612: Deterministic Models in Operations Research I
3. STOR 641: Stochastic Models in Operations Research I

In addition, the students must take **seven electives** and **one presentation course** (STOR 790). At least four of the seven electives courses must be STOR courses at least two of which must be at level 600 and above. **The student may select further electives from other departments such as Business School at UNC, Business School at Duke, Computer Science, Economics, etc.** However, all course selections must be approved by the graduate advisor.

Recommended Electives:

- STOR 456: Time Series, Forecasting, Data mining
- STOR 565: Machine Learning
- STOR 614: Deterministic Models in Operations Research II
- STOR 642: Stochastic Models in Operations Research II
- STOR 664: Applied Statistics I
- STOR 665: Applied Statistics II
- STOR 743: Stochastic Models in Operations Research III
- STOR 762: Discrete Event Simulation
- STOR 712: Non-linear Programming
- STOR 722: Integer Programming
- STOR 724: Networks
- STOR 705: Practicum
- MBA 748A: Marketing Analytics
- BA 830: Operations Management I
- BA 832: Operations management II
- BA 837: Decision Making with Spreadsheets
- Quantitative Methods in Marketing
- STOR 992: Masters Paper
- COMP 790-90: Data Mining – Concepts, Algorithms, Applications
- ECON 871: Time Series and Forecasting
- ECON 710: Advanced Microeconomic Theory I

Total hours required for degree: 31

Computational Social Sciences for Undergraduate and Graduate Students

Andrew Cabaniss

March 25, 2014

Quantitative and computational methods have emerged as essential tools for researchers across the social sciences. Complex social systems have long merited qualitative study, but powerful computational tools have begun to enable quantitative hypothesis testing and modeling on a routine basis. Such methods extend the domain of application for theories developed in fields like archaeology and sociology, which increases the impact of research and the desirability of students for relevant jobs. They ease interdisciplinary communication and can strengthen collaborations between diverse groups of scholars and students.

The current undergraduate and graduate curriculum in most departments does not address the increasing importance of quantitative and computational methods in the active scholarship of the social sciences. Moving beyond the basic statistical frameworks of the past, UNC Arts and Sciences should strive to prepare students to develop and use cutting edge tools for analyzing and understanding complex social systems. Such methods are rarely taught directly to undergraduates despite their wide-ranging applicability and ease of use.

Courses exist that cover statistical methods aimed at the social sciences, but these primarily allow students to engage with the existing scholarship rather than forge ahead in new directions or engage with the most recent progress in these fields. Domain-specific quantitative seminars exist in psychology, economics, and sociology, but provide a limited toolkit in comparison with the social sciences as a whole. To supplement these existing courses, I propose a two-semester course sequence, parallel in intent with applied mathematics and courses such as "Linear Algebra with Applications." It would specifically target social science applications and provide a wide range of people with the basic skill sets to understand and conduct research in the emerging frontiers of the social sciences. A two-semester course could serve students continuing to specialize in their majors without disrupting their courseload through years of math courses, and would enable further studies in topics of interest through the existing array of statistics courses scatted across the various departments.

In this proposal, I first review the primary topics covered in the applicable courses at

UNC for which I was able to acquire syllabi. I then examine syllabi for other courses in computational social science at other universities, and propose a set of topics applicable to active and progressing research across these fields. I then lay out a framework for a two-semester computational social science course on this basis.

Computational Social Science Methods Courses at UNC

I examined the course syllabus for eight courses at UNC that cover statistical or quantitative methods available or aimed at students of the social sciences (see table 1). I noted the topics covered in each course of interest, both those important for previous as well as current scholarship. The topics included statistical inference, linear regression, basic probability distributions, model selection, hypothesis testing, Markov chains, network theory, game theory, information theory, and agent based modeling.

Table 1: Computational and Quantitative Social Science Methods Courses at UNC

Course number	Course name
STOR 455	Statistical Methods I
STOR 456	Statistical Methods II
STOR 155	Introductory Statistics
STOR 215	Introduction to the Decision Sciences
PSYC 210	Statistical Principles in Psychological Research
PSYC 215	Statistical Principles in Psychological Research
ECON 400	Elementary Statistics

Across these courses, UNC provides many opportunities to learn elementary quantitative methods such as linear regression, statistical inference, and common distributions. Statistics, psychology, and economics each provide at least one course that covers these topics. Other methods are not readily covered, however, across these courses. None of these courses teaches information theory. Game theory, Markov chains, and networks are only covered in a single introductory course (STOR 215). These have become increasingly important in recent research, and have dominated the discourse in recent quantitative and computational social science studies. Further, these methods have not been applied in a wide range of contexts, creating lucrative opportunities for those who are able to use them. This would allow them to interact with other researchers on the cutting edge of the social sciences. In short, UNC's current offerings allow basic quantitative frameworks, but lack an introduction to complex statistical and computational methods for the social sciences.

Computational Social Science Methods at Other Universities

I contrasted four courses at four universities with the offerings at UNC (see table 2). These were selected through Complexity Explorer, an online tool for complexity and complex science education managed by the Santa Fe Institute, as representative courses aimed at undergraduates and graduate students taught by some of the leading figures in complex social systems.[1]

Table 2: Computational and Quantitative Social Science Courses at Other Universities

Course name	Course Instructor, University
Large Scale Social Phenomena	Simon DeDeo, Indiana University
Inference, Models and Simulation for Complex Systems	Aaron Clauset, CU Boulder
Introduction to Scientific Modeling	Stephanie Forrest, UNM
Networks and Complexity in Social Systems	Duncan Watts, Columbia

These courses have a stronger emphasis on networks, game theory, Markov chains, and Bayesian inference, topics covered by three of the four syllabi. Notably, they provide significantly less information on model selection and hypothesis testing, both critical for justifying modeling assumptions and decisions. Only one course provides any material on information theory or signaling. These classes present a complementary suite of material to that currently taught at Carolina, providing introductions to method and theory for the less historically important of the current major analytical frameworks.

In these courses, material is often taught conceptually and methodologically rather than as a derivation *ex nihilo* which requires high-level mathematics. Emphasis on applications rather than technical detail combined with the use of computational methods continues the trend of working over generative knowledge. In at least one case (DeDeo), a separate teaching assistant helped students with computational matters while the professor taught methods and theory, allowing the students in the social sciences who have not had any computational background to learn the basics of scientific and mathematical computing side-by-side with relevant applications.

A New Course for Computational Social Science

In tackling complicated problems foundations are critical to arguments and ideas. In quantitative and computational social science, a fundamental understanding of probability and

[1] http://www.complexityexplorer.org/

inference provides the necessary background to tackle higher-order, complex problems. Complex data allow us to examine new hypotheses, structures, and behaviors, but only provided that we have adequate analytical tools. At present, UNC provides many opportunities for learning the basics, but few for learning applied or complicated methods without extensive training in mathematics and statistics, despite the importance of these methods in the current social sciences. Other state universities have successfully implemented and taught courses for precisely these theories and methods aimed at social scientists, although they often lack the underpinnings of basic probability and inference. Combining these approaches, I would like to define a rough layout for a two-part course sequence that corresponds in its halves to the two approaches displayed above.

The first semester (table 3) would condense the theoretical background in probability theory that most courses provide with Bayesian inference, model selection, and the basic probability distributions with their useful properties. This would proceed at a higher level than STOR 155 but with a similar broad scope. The emphasis would be on problem solving using these methods, in part to contextualize learning but also to facilitate the creation of a comprehensible toolbox for students to use in future work. It would familiarize students with the computational background to analyze data in their own field and the theoretical underpinnings to determine the appropriate situations to use diverse tools.

The second semester (table 4) would build upon this toolbox and expand it with theories that are often probabilistically founded. Markov chains, networks, information theory, and game theory would be treated, with introductions to theoretical backgrounds, analytical tools, and case studies from recent years across the social sciences. This combination of topics would enable students to pursue their own research, engage with current discourses in a variety of social disciplines, and provide the motivation as well as the technical background to continue on into more advanced methods courses they found appealing, either within or across disciplines.

Given the wide range of people who would be interested in such a course, I would suggest a model that allowed a professor or group of professors to lecture on theory, method, and notable application, while a recitation or lab section would focus on problem sets, data analysis, and computational problems with small groups of students. This course would also work well in a flipped classroom setting, given the stronger emphasis on in-class application and the easier means to team-teach a class by providing online content from a variety of instructors and sources. Such a course would provide more analytical tools for students to use in their work and studies, fill gaps in the current offerings available to Carolina students, and facilitate new discoveries on the emerging frontiers of the social sciences.

Table 3: First Semester: Probability and Inference

- Probability
  - Axioms of Probability
  - Independence, Joint Probabilities, Conditional Probabilities, and Bayes Theorem
  - Moments and Properties
  - Common Distributions and Interrelations
  - Calculating and manipulating probabilities and their properties using SAS, STATA, R, or Python
    * Calculating probabilities given events
    * Distributions: Construction and Sampling
    * Numerical integration methods for probability distributions
    * Plotting and presenting results
- Inference and Modeling
  - Inference and Maximum Likelihood;
  - First Order Inference of Parameters and Confidence
  - Second Order Inference and Model Selection
  - Implementing and using parameter estimation techniques in computer software packages from real data; computational techniques for handling second and first order inference problems
    * Bootstrapping and low-data situations
    * Linear regression
    * Evidence and likelihood assessment through gradient and numerical integration approaches
    * Plotting and presenting results

Table 4: Second Semester: Theories of Content, Strategy, and Structure

- Information
  - Symbols, Information, and Entropy
  - Maximum entropy and information as scientific knowledge
  - Markov Chains
  - Generative Languages
  - Computational techniques for dealing with information and language
    - Assessing information content and mutual information using naive and bootstrap estimators
    - Modeling and fitting Markov chains using maximum likelihood gradient methods
    - Maximum entropy methods for inference under constraints
- Games
  - Signaling and Cooperative Games, Introduction to Utility/Payoff and Decisions
  - Non-cooperative Games, Strategy, Equilibrium Solutions
  - Simulation of games, gradient methods for strategy assessment and payoff calculation
- Networks
  - Nodes and Edges, Basic Properties and Types of Networks, Random Graphs
  - Clustering, Communities, Modularity, and Random Graphs w/ Communities
  - Signaling and Games on Networks
  - Software and packages for handling networks of different types and sizes;
    - Calculating properties of networks
    - Generating random graphs
    - Community detection
    - Simulation of network games
    - Network model selection