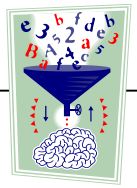
MMC 6936 Big Data Seminar

Spring 2016 / Section 1E86 / Periods 8-10 Mondays, Weimer 3020 / 3 credits



Associate Professor Norman P. Lewis, Ph.D.

Office: Weimer 3052
Office hours: Periods 4-5 (10:40a to 12:35p) Mondays & periods 8-9 (3p to 5p) Tuesdays
Phone: 352-392-5137 (I respond within 24 hours Monday through Friday)
E-mail: nplewis@ufl.edu (I respond within 24 hours Monday through Friday)
About me: See my website or follow me (@bikeprof) on Twitter

"Office hours" are times I dedicate to being in the office (unless called away by my bosses) and are first-come, first-served. Or, make an appointment. Or, drop in if the door is open.

ABOUT THE COURSE

Course Description

This course unpacks the phenomenon of big data and how large data sets and sophisticated algorithms are changing society in general and mass communication in particular. The course also enables you acquire and evaluate a large data set for a research paper (although no experience in data or statistics is presumed or required).

Course Objectives

The course is intended to enable you to:

- Distinguish "big data" from "data."
- Identify the key benefits and dangers of big data for both individuals and society.
- Debate the ethical tension between retaining privacy and withholding information.
- Analyze how datification is altering the social compact.
- Evaluate how big data is changing the practice of mass communication and media.
- Critique research using big data as well as research about the phenomenon.
- Produce a research paper using a large data set.

Required Textbooks

1. Tucker, P. (2014). *The naked future: What happens in a world that anticipates your every move?* New York, NY: Current. ISBN: 978-1591845867 (online: Amazon or Barnes & Noble). This popular-press book sees predictive analytics in a hopeful light.
2. Kitchin, R. (2014). *The data revolution: Big data, open data, data infrastructures and their consequences*. Thousand Oaks, CA: Sage. ISBN: 978-1446287484 (online: Amazon or Barnes & Noble). This book offers a broad overview of the subject from an academic perspective.

Supplemental Textbook (not required)
Borgman, C.L. (2015). *Big data, little data, no data: Scholarship in the networked world*. Cambridge, MA: MIT Press. ISBN: 978-0262028561. Although not focused on big data, this scholarly book is an excellent explication of data, regardless of size.

Course Readings
Because the two assigned texts are overviews, both are to be read by our second class meeting, on January 25. (UF is closed on January 18 in honor of Martin Luther King Jr.)

The Kitchin text is the only one I have found that addresses big data from an academic perspective. It is a bit dry but useful. The Tucker book is a popular press book and thus much more enjoyable to read.

In addition to the two textbooks, readings are assigned for each week. Those readings are available from the course page on Canvas.

Some of the readings are research articles from peer-reviewed journals. Because the field of big data is so new, relatively few good research articles on big data exist. I have selected some of the better ones from what I could find by searching library databases as well as *Big Data and Society* journal, which started in 2014. Those readings are online on the Canvas website as PDFs you can download.

Because so few research articles exist – and because the field of big data is changing so fast – a number of journalistic articles are also included in your readings. Most of these are from the financial press, primarily the *Wall Street Journal*. Because *WSJ* hides its work behind a paywall, I have saved the articles as PDFs and posted them to Canvas. Please preserve those *WSJ* articles for your own use and do not share them with others or post them online.

Where articles are not hidden behind a paywall, the Canvas pages offer hyperlinks to websites. Among these are the *New York Times*, which limits how many articles you can view free per month. As a UF student, you can escape those limitations by getting a free subscription through your web browser or a smartphone app. If you don't already have a free NYT account, click here for instructions.

Each week's readings should be read ahead of class. I will call on you in class with the expectation that you have read all the assignments. However, circumstances may arise that keep you from completing the reading. In that case, you get one "free pass" during the semester. To claim it, email me ahead of class so I won't call on you. But do come to class.

ASSIGNMENTS AND GRADING

Tentative Schedule

	Date	Topic	Details
1	Jan. 11	Overview	Understanding data and knowledge; criteria and tools that distinguish "big data" from "data."
	Jan. 18	(No class; MLK Jr. Day)	
2	Jan. 25	Research About Big Data	What does early research done on the practice and potential of big data tell us about both the phenomenon and the research itself?
3	Feb. 1	Research Using Big Data	Are studies using big data as a tool new and enlightening or just another verse to the same research song?
4	Feb. 8	MassMine 1	Class 1 of 3 exploring how to do your research paper with MassMine. Bring your laptop.
5	Feb. 15	MassMine 2	Class 2 of 3 exploring how to do your research paper with MassMine. Bring your laptop.
6	Feb. 22	MassMine 3	Class 3 of 3 exploring how to do your research paper with MassMine. Bring your laptop.
	Feb. 29	(No class; spring break)	
7	March 7	Individual meetings	In lieu of class, we will hold individual meetings to discuss your plans for your paper.
8	March 14	Advertising	Data mining and predictive analytics – or why mass advertising is fading while Facebook and Google are filthy rich.
9	March 21	Mass vs. Social	Are social networks shaping or replacing mass media? Can dual screens, addressable TV, and online analytics save mass media?
10	March 28	Datification	From Fitbits to smart homes to pollution detectors to dating sites, we measure more each day. The more, the merrier?
11	April 4	Governing	Do you want government to know which college students are more likely to drop out or who is more likely to commit crime?
12	April 11	Social Good	Should companies use big data to raise insurance rates for unhealthy people or detect slackers at work?
13	April 18	Ethics & Privacy	Is your data be your private property or do you have an obligation to share data that would benefit others?

Assignment Weighting

Class presentation 20%
Essay final exam 20%
Research paper 60%

Grading Scale

		B+	89 to 87	C+	79 to 77	D+	69 to 67
A	100 to 93	B	86 to 83	C	76 to 73	D	66 to 63
A-	92 to 90	B-	82 to 80	C-	72 to 70	D-	62 to 60

Consult the UF graduate school catalog for details about the grading policy.

Class Presentation (20%)
The first day of class, you will pick a week to lead the class is a teaching discussion of about 30 to 45 minutes. Your task is to (a) supplement the assigned readings and (b) engage the class in a discussion suitable for a graduate seminar.

In general:

- An excellent (A) presentation is a polished educational presentation that offers fresh material and insight while engaging the audience in a meaningful way.
- A very good (B) presentation offers perhaps more synthesis than fresh material but is still valuable while engaging the audience.
- An unsatisfactory presentation offers little more than a summary of the assigned readings or is a "death by PowerPoint" too reliant on reading a script or notes.

The scoring rubric assigns points (total: 100) is:

Points	Description	Result
Presentation Materials (65 points)		
5	Presentation roughly split between assigned readings and supplemental material found by the presenter.	
10	Assigned readings are synthesized (not summarized) through contextualization, comparison, and analysis.	
10	New material extends the assigned readings by offering fresher or different perspectives.	
15	Materials involver as much visual as words, and visuals (pictures, graphics, videos, etc.) are illustrative or explanatory.	
10	If PowerPoint used, words are in 32-point type and brief, or no more than about 30 words per slide.	
10	Materials are clear (easy to follow with terms defined without jargon) and concise (to be point; without padding).	
5	English is easily understood: without significant errors in syntax grammar, spelling or punctuation.	
Presentation Mechanics (20 points)		
10	Presenter smiles, speaks clearly and audibly, and uses eye contact effectively.	
10	Poised speech style reflects superior preparation rather than script memorization or more-than-passing references to notes.	
Classroom Engagement (15 points)		
5	Roughly a third of presentation time is spent engaging classmates.	
5	Engagement is built on thoughtful, open-ended questions or hands-on approaches to involve peers.	
5	Presenter invites questions and is prepared to answer them.	

Essay Final Exam (20%)

At the conclusion of the semester, you will be assigned an essay exam with 2 or 3 questions.

1. The questions will be derived from the course objectives in this syllabus. For example: "Debate the ethical tension between retaining privacy and withholding information, and draw from an ethical philosophy discussed in class to resolve it."
2. Compose your answers using a word processing program such as Microsoft Word.
3. Upload the answers as a file to an assignment on the Canvas website that will also use the Turnitin plagiarism-detection software.
4. The exam will be open book so you can cite any of the material used in class.
5. The exam must be taken alone and without any assistance, virtual or physical.
6. All will take the exam simultaneously. The class will choose the time and date.
7. At the agreed-upon time, the questions will be visible on Canvas and you will have four hours (same time a doctoral student has to complete a qualifying exam) to upload your answers.
8. Answers will be judged similarly for master's as for doctoral students.

The scoring rubric for the essay exam is:

	100-90 Excellent	**89-80 Good**	**Less than 80 Unsatisfactory**	**Result**
Content (70%)	Essay answers all questions clearly, accurately, and succinctly. Answers demonstrate superior understanding about big data by offering more synthesis than summary in drawing from multiple sources.	Essay answers all questions clearly and accurately. Answers reveal a competent understanding of big data based on a fair summary of multiple sources.	Essay answers all questions accurately if not always clearly or succinctly. Answers reveal incomplete understanding of big data, and may draw from an insufficient range and number of sources.	
Reasoning (15%)	Essay offers superior reasoning based on persuasive evidence without "I think / feel / believe" statements.	Essay offers reasoning based on evidence without "I think / feel believe" statements.	Essay offers reasoning too often based on personal opinion rather than evidence.	
Mechanics (20%)	Writing is engaging and lucid with precise word choice and clear syntax. Any errors in punctuation, spelling, or grammar are few and generally immaterial.	Writing is understandable. A few errors in word choice, syntax or writing mechanics may slow readability but do not impede comprehension.	Writing is too often difficult to decipher because of faulty reasoning, garbled syntax, or mechanical errors.	

Research Paper (60%)

This research paper will draw from a large set of original data drawn from MassMine.

1. Write the paper in the style of a research paper for a peer-reviewed journal, starting with an abstract of 100 words that summarizing the findings of the paper and then structured according to the standard five sections of a research paper:
    a. Introduction: Define the problem to be solved (the "so what?"). Identify the gap in academic literature that the study will fill.
    b. Literature review: Explicate theory base for study. Define and explicate all concepts and variables that lead to hypotheses or research questions.
    c. Method: Defend the method as a suitable way to answer or address the H's and RQs. Detail and justify the research design. Operationalize all variables.
    d. Findings: Apply the data gathered and analyzed through statistical testing to the H's and RQs. Analyze each result and its practical significance.
    e. Conclusion: Synthesize findings to parse out both discrete and big-picture meaning. Identify limitations. Suggest future research. Finish with a final paragraph "selling" the study by answering the "so what?" question.
2. The paper should be 4,000 to 5,000 words excluding references, or about 20 to 25 double-spaced pages in total.
3. Follow APA style, which in its style book offers annotated examples of research papers. An abbreviated summary of APA style is on the Canvas website (although the full guide book is much better).
4. You may choose to work alone or collaboratively on this paper.
    a. If you work alone, solo papers from master's students will be graded more leniently.
    b. If you collaborate with someone else in this class, each of you will be required to email me answers to these questions:
        i. What evidence can document your contributions to the paper?
        ii. What knowledge or skill did you bring to the project?
        iii. What knowledge or skill did you learn from your co-author?
5. The grading rubric is:
    a. Grade A: Paper is of sufficient quality that it is ready for acceptance to a regional research conference.
    b. Grade A-: Paper is of sufficient quality that with just a few relatively minor revisions, it would be ready for acceptance to a regional research conference.
    c. Grade B+ / B / B-: Paper could be made ready for acceptance to a regional research conference with more than a few, but not a lot of, revisions.
    d. Less than a B-: Paper needs a lot of revisions before it would be ready for acceptance to a regional research conference.

MORE IMPORTANT DETAILS

Attendance

Punctual attendance is expected of graduate students unless an absence is allowed under UF policy. If you are going to be absent, let me know ahead of time.

Academic Integrity

UF students live by an honor code that prohibits academic dishonesty such as (but not limited to) cheating, plagiarism, fabrication, engaging in unauthorized collaboration, reusing your master's thesis or a paper from another class, writing a similar paper for two classes, drawing too heavily on another's work for your own, and having someone else write your paper.

Be aware of the UF graduate school academic honesty policy as well the one in the college's Doctoral Handbook. Students have an affirmative obligation to know what is in the handbook and to abide by it. The handbook includes a detailed description of plagiarism, copies of which are available in Chinese, Korean, Mandarin, Portuguese, and Spanish. If you are unsure of citation rules or what requires attribution, ask me *before* turning in a paper. Ignorance is not an excuse.

My default practice for an academic integrity violation is a failing grade for the course and to recommend the student be removed from the graduate program.

Students with Disabilities

If you would benefit from disability-related accommodations, contact the Disability Resource Center as early in the semester as possible. The center will provide documentation so appropriate accommodations can be made. The center is in Reid Hall, 392-8565.

Help with Coping

The UF Counseling and Wellness Center is a terrific, free resource for any student who could use help managing stress or coping with life. The center, at 3190 Radio Road on campus, is open for appointments and emergency walk-ins from 8 a.m. to 5 p.m. Monday through Friday. To make an appointment or receive after-hours assistance, call 352-392-1575.

The UF Police can be reached at 392-1111 or, in an emergency, by dialing 911.

Course Evaluations

Students are expected to provide feedback on the quality of instruction in this course by completing online evaluations. You will be notified by email when the evaluations are open, near the end of the semester. Summary results are available to you and the general public.