



Stony Brook University
Department of Technology and Society
College of Engineering and Applied Sciences

EMP 532 – Big Data Systems for Technology Management

Important Note: All times and deadlines in the syllabus are based on UTC -4 (time in New York). Every effort will be made to avoid changing the course schedule, but the possibility exists that unforeseen events will make syllabus changes necessary. It is your responsibility to check Blackboard for corrections or updates to the syllabus. Any changes will be clearly noted in course announcements or through Stony Brook email.

Contents

Instructor Information.....	2
Course Description:	2
Textbook:	2
Helpful Books:	2
Course Delivery Mode and Structure:	3
How We Will Communicate:.....	3
Technical Requirements:	3
Technical Assistance:.....	4
Learning objectives:.....	4
How to Succeed in this Course:	4
Course Schedule.....	4
Grading, Attendance, and Late Work Policies.....	5
Assessment and Grading:.....	5
Viewing Grades on Blackboard	5
Letter Grades:.....	6
Attendance Policy:	6
Late Work Policy:	6
Other Information	6
Course and University Policies.....	7
Student Accessibility Support Center Statement:	7

Academic Integrity Statement:	7
Critical Incident Management:	7
Understand When You May Drop This Course:.....	8
Incomplete Policy:	8
Course Materials and Copyright Statement:	8
Online Communication Guidelines and Learning Resources:.....	8
Online Etiquette:	8
Online Classes Require Better Communication	8
My Role as the Instructor:.....	9
Student Resources:	9

Instructor Information

INSTRUCTOR: Dr. Thomas S. Woodson
 E-MAIL: thomas.woodson@stonybrook.edu
 OFFICE PHONE: (631) 632-9974
 OFFICE HOURS: By appointment through email

Course Description:

Course title: Big Data Systems for Technology Management
 Course catalog # and section: EMP 532.01
 Credit hours: 3
 Prerequisites: Must be a Master's student in Technology and Society or receive permission from the instructor

Welcome. I am excited to teach EST 532. I've taught this course several times and each year it is an enriching experience for both me and the students. Over the past two decades, the amount of data that is generated has grown exponentially and there is an increasing need to analyze all this data. This class will introduce students to the statistical software R, data analysis, text mining, and big data analyses.

Textbook:

Data and Goliath: The Hidden Battles to Collect Your Data and Control your World
 Bruce Schneier

Helpful Books:

- An Introduction to Applied Multivariate Analysis with R by Brian Everitt and Torsten Hothorn
- A Beginner's Guide to R, by Alain Zuur, Elena Ieno, and Erik Meesters, Springer 2009.
- R in a Nutshell: A Desktop Quick Reference, by Joseph Adler, O'Reilly 2010.
- An R and S-PLUS Companion to Multivariate Analysis, by Brian Everitt, Springer 2005.

- The R Book, by Michael Crawley, Wiley 2007.
- Big Data Analytics with R, by Simon Walkowiak

Course Delivery Mode and Structure:

This is a synchronous online course, delivered in the Blackboard learning management system (LMS). Students must be mindful of all course expectations, deliverables and due dates, especially because the online portion of the course requires significant time management. All assignments and course interactions will utilize internet technologies. See “Technical Requirements” section for more information. In Blackboard, you will access online lessons, course materials, and resources. EACH WEEK HAS 1 major LEARNING MODULES. All assignments should be submitted the deadline of UTC -4 (New York Time).

How We Will Communicate:

Course-related questions should be posted in the General Questions Forum in the course Discussion board. For personal/private issues, email me directly. If you use Blackboard’s **email tool** from the course site, it will automatically include your full name, course name and section when you send me an email. **Please allow between 24-48 hours for an email reply.** Your Stony Brook University email must be used for all University-related communications. You must have an active Stony Brook University email account and access to the Internet. All instructor correspondence will be sent to your SBU email account. **Plan on checking your SBU email account regularly for course-related messages.** To log in to Stony Brook Google Mail, go to <http://www.stonybrook.edu/mycloud> and sign in with your NetID and password.

Regular announcements will be sent from Blackboard. These will be posted in the course site and may or may not be sent by email.

Regular communication is essential in online classes. Logging in once a day, checking the discussion board and participating with your colleagues ensures that you are able to remain an active member of the class and earn full points for participation.

Technical Requirements:

This course uses Blackboard for the facilitation of communications between faculty and students, submission of assignments, and posting of grades and feedback. The Blackboard course site can be accessed at <https://blackboard.stonybrook.edu>

If you are unsure of your NetID, visit <https://it.stonybrook.edu/help/kb/finding-your-netid-and-password> for more information. You are responsible for having a reliable computer and Internet connection throughout the term. **Caution!** You will be at a disadvantage if you attempt to complete all coursework on a smart phone or tablet. It may not be possible to submit the files required for your homework assignments.

Students should be able to use email, a word processor, spreadsheet program, and presentation software to complete this course successfully.

The following list details a minimum recommended computer set-up and the software packages you will need to have access to, and be able to use:

- PC with Windows 8 or higher (we recommend a 3-year Warranty)

- Macintosh with OS 10.11 or higher (we recommend a 3-year Warranty)
- Intel Core i5 or higher
- 250 GB Hard Drive
- 8 GB RAM
- Latest version of Chrome, Firefox or Explorer; Mac users may use Chrome or Firefox. (A complete list of supported browsers and operating systems can be found on the My Institution page when you log in to Blackboard.)
- High speed internet connection
- Webcam
- Printer
- Word processing software (Microsoft Word, Pages, etc.)
- Speakers (either internal or external) or headphones
- Ability to download and install free software applications and plug-ins (note: you must have administrator access to install applications and plug-ins).
- Adobe Flash player with the latest update is crucial for playing multiple videos throughout the course

Technical Assistance:

If you need technical assistance at any time during the course or to report a problem with Blackboard you can:

- Phone: 631-632-9800 (client support, Wi-Fi, software and hardware)
- Submit a help request ticket: <https://it.stonybrook.edu/services/itsm>
- If you are on campus, visit the Walk-Up Tech Support Station in the Educational Communications Center (ECC) building.

Learning objectives:

Upon completion of the course, you will be able to:

1. Make plots using the GGplot package in R
2. Do linear regression and logistic regression analyses
3. Cluster data with techniques like k-means and hierarchal clustering
4. Conduct basic text mining analyses
5. Discuss basic principles of big data analysis

How to Succeed in this Course:

- Complete all assigned readings in the course
- Practice R everyday
- Write your own code instead of just “copy and pasting” what I write.
- Ask for help if you have errors.
- You should expect to spend at least 10-12 hours per week outside of class working on the homework and practicing R.

Course Schedule

Course calendar subject to change

Lecture	Tuesday	Thursday	Topic	Major Assignments Due
Week 1	25-Jan	27-Jan	Intro to Data/Intro to R	
Week 2	1-Feb	3-Feb	Intro to R/GGPlot	
Week 3	8-Feb	10-Feb	GGPlot/Practice	
Week 4	15-Feb	17-Feb	R Test 1/Book Test 1 Mapping	
Week 5	22-Feb	24-Feb	Mapping/Clustering	Project 1: GG Plot Due (Friday)
Week 6	1-Mar	3-Mar	Clustering/Regression	
Week 7	8-Mar	10-Mar	Regression/Logit	
	15-Mar	17-Mar	No class, Spring Break	
Week 8	22-Mar	24-Mar	Book Test 2/Logit	
Week 9	29-Mar	31-Mar	R Test 2/Work on projects	
Week 10	5-Apr	7-Apr	Text Mining	Project 2: Data analysis Due (Friday)
Week 11	12-Apr	14-Apr	Data and Goliath Ch 12-14 /Text Mining	
Week 12	19-Apr	21-Apr	Data and Goliath Ch 15-16/ Book Quiz 3/Text Mining	Project 3-Canceled
Week 13	26-Apr	28-Apr	Big Data/Research Day	
Week 14	3-May	5-May	Big Data/Project	Project 4 Due

Grading, Attendance, and Late Work Policies*Assessment and Grading:*

The main assignments for this class are 3 class projects. In addition to the projects the class will feature discussion boards, and tests.

Viewing Grades on Blackboard:

Points and feedback for graded activities will be posted to the My Grades tab in the Tools area of Blackboard. In most cases, discussion boards, VoiceThreads, Journals will be graded within 48 hours of closing each week. Submitted projects will be graded within 2 weeks of being submitted.

In this course, you will be assessed on the following:

Activity/Assignment	Points
Book quizzes/Discussion boards	25
Projects	50

R Quizzes	15
Participation/other assignments	10
Total	100

Letter Grades:

Final grades assigned for this course will be based on the percentage of total points earned and are assigned as follows:

Letter Grade	Points	GPA/Points	Performance
A	94-100	4.0	Excellent work
A-	90-93	3.7	Nearly excellent work
B+	87-89	3.3	Very good work
B	83-86	3.0	Good work
B-	80-82	2.7	Mostly good work
C+	77-79	2.3	Above average work
C	73-76	2.0	Average work
C-	70-72	1.7	Mostly average work
D+	65-69	1.3	Below average work
D	60-34	1.0	Poor work
F	Less than 60	0.0	Failing work

Attendance Policy: In order to succeed in this class, you must attend class. If you regularly miss class, it will negatively impact your participation points for the course.

Late Work Policy: Turn in your work by the due date in order to receive full credit. After the deadline you have the option to turn in work, but you will lose points. Late work will only be accepted with full credit given university sanctioned absences (ie. Documented illness, or documented family emergencies)

Other Information

Discussion Board Grading Rubric

Interpretation Points	Quality of posts	Frequency	Mechanics
Exemplary 5, full points or higher	The comment is accurate, relevant, properly attributed and evidence-based as well as original and well written. Adds substantial teaching presence to the course and stimulates additional thought about the issue under discussion. Collegial and friendly tone.	Participates steadily throughout the week (or module) and responds to instructor and/or peers on or before deadline.	Free of spelling and grammatical errors.

Accomplished 3 to 4	The comment lacks at least one of the above qualities, but is above average in quality. Makes a significant contribution to our understanding of the issue being discussed.		One or two minor errors.
Developing 2 to 3	The comment lacks two or three of the required qualities. Comments which are based solely upon personal opinion or personal experience often fall within this category.	Few posts. Posts are bunched into one or two days at the end of the week.	
Needs work 1 to 2	The comment presents little or no new information. However, may provide social presence and contribute to a collegial atmosphere.	Few posts. Deadlines are not met.	Multiple spelling and grammar errors or inappropriate.

Course and University Policies

Student Accessibility Support Center Statement:

If you have a physical, psychological, medical, or learning disability that may impact your course work, please contact the Student Accessibility Support Center, 128 ECC Building, (631) 632-6748, or via e-mail at: sasc@stonybrook.edu. They will determine with you what accommodations are necessary and appropriate. All information and documentation is confidential.

Academic Integrity Statement:

Each student must pursue his or her academic goals honestly and be personally accountable for all submitted work. Representing another person's work as your own is always wrong. Faculty are required to report any suspected instances of academic dishonesty to the Academic Judiciary. Faculty in the Health Sciences Center (School of Health Technology & Management, Nursing, Social Welfare, Dental Medicine) and School of Medicine are required to follow their school-specific procedures. For more comprehensive information on academic integrity, including categories of academic dishonesty please refer to the academic judiciary website at http://www.stonybrook.edu/commcms/academic_integrity/index.html.

Important Note: Any form of academic dishonesty, including cheating and plagiarism, will be reported to the Academic Judiciary.

Critical Incident Management:

Stony Brook University expects students to respect the rights, privileges, and property of other people. Faculty are required to report to the Office of University Community Standards any disruptive behavior that interrupts their ability to teach, compromises the safety of the learning environment, or inhibits students' ability to learn. Faculty in the HSC Schools and the School of Medicine are required to follow their school-specific procedures. Further information about most academic matters can be found in the Undergraduate Bulletin, the Undergraduate Class Schedule, and the Faculty-Employee Handbook.

Understand When You May Drop This Course:

It is the student's responsibility to understand when they need to consider withdrawing from a course. Refer to the Stony Brook Academic Schedule for dates and deadlines for registration:

Incomplete Policy:

Under emergency/special circumstances, students may petition for an incomplete grade. Circumstances must be documented and significant enough to merit an incomplete. If you need to request an incomplete for this course, contact me for approval as far in advance as possible.

Course Materials and Copyright Statement:

Course material accessed from Blackboard, SB Connect, SB Capture or a Stony Brook Course website is for the exclusive use of students who are currently enrolled in the course. Content from these systems cannot be reused or distributed without written permission of the instructor and/or the copyright holder. Duplication of materials protected by copyright, without permission of the copyright holder is a violation of the Federal copyright law, as well as a violation of Stony Brook's Academic Integrity.

Online Communication Guidelines and Learning Resources:

Maintain Professional Conduct Both in the Classroom and Online: The classroom is a professional environment where academic debate and learning take place. I will make every effort to make this environment safe for you to share your opinions, ideas, and beliefs. In return, you are expected to respect the opinions, ideas, and beliefs of other students—both in the face-to-face classroom and online communication. Students have the right and privilege to learn in the class, free from harassment and disruption. The course follows the standards set in the Student Code of Conduct, and students are subject to disciplinary action for violation of that code. If your behavior does not follow the course etiquette standards stated below, the grade you receive for a posting may suffer. I reserve the right to remove any discussion messages that display inappropriate language or content.

Online Etiquette:

- Offensive language or rudeness will not be tolerated. Discuss ideas, not the person.
- Avoid cluttering your messages with excessive emphasis (stars, arrows, exclamations).
- If you are responding to a message, include the relevant part of the original message in your reply, or refer to the original post to avoid confusion;
- Be specific and clear, especially when asking questions.
- Use standard punctuation and capitalization. Using all UPPERCASE characters gives the appearance of shouting and makes the message less legible;
- Remember that not all readers have English as their native language, so make allowances for possible misunderstandings and unintended discourtesies.

Online Classes Require Better Communication:

It is important to remember that we will not have the non-verbal cues that occur in a face-to-face classroom. I cannot see the confused, frustrated, or unhappy expressions on your face if you encounter problems. You **MUST** communicate with me so that I can help. To make the experience go smoothly, remember that you're responsible for initiating more contact, and being direct, persistent, and vocal when you don't understand something.

My Role as the Instructor:

As the instructor, I will serve as a “guide” in our online classroom. While I will not respond to every post, I will read what is posted, and reply when necessary. Expect instructor posts in the following situations:

- To assist each of you when it comes to making connections between discussion, lectures, and reading material.
- To fill in important things that may have been missed.
- To re-direct discussion when it gets “out of hand.”
- To point out key points or to identify valuable posts.

Student Resources:

- U.S. Federal Statistics: <http://fedstats.sites.usa.gov/>
- Create data visualization: <http://www.creativebloq.com/design-tools/data-visualization-712402>
- DASL (The Data and Story Library): <http://lib.stat.cmu.edu/DASL>
- The Electronic Encyclopedia of Statistical Examples and Exercises <http://www.macmillanhighered.com/catalog/static/whf/eesee/eesee.html>
- JASA (Journal of the American Statistical Association) Data Archive <http://lib.stat.cmu.edu/jasadata/>
- JSE (Journal of Statistics Education) Data Archive http://www.amstat.org/publications/jse/jse_data_archive.html
- Statlib-Datasets Archive <http://lib.stat.cmu.edu/datasets/>
- University of California, Los Angeles Case Studies <http://www.stat.ucla.edu/cases/>
- U.S. Bureau of Labor Statistics <http://stats.bls.gov>
- U.S. Census Bureau <http://www.census.gov>
- Stats in the news, from George Mason University: <http://www.stats.org/>

Online statistics textbooks and software:

- Computing for Data Analysis <https://www.coursera.org/course/compdata>
- Data Analysis <https://www.coursera.org/course/dataanalysis>
- Rice virtual lab in statistics <http://onlinestatbook.com/rvls.html>
- SISA simple interactive statistical analysis <http://www.quantitativeskills.com/sisa/>

Online resources for R:

- <http://r-statistics.co/Linear-Regression.html>
- The main R project site:
www.r-project.org
- An R online textbook
[Kickstarting R: http://cran.r-project.org/doc/contrib/Lemon-kickstart/](http://cran.r-project.org/doc/contrib/Lemon-kickstart/)
- Website for the Sarkar book:
<http://lmdvr.r-forge.r-project.org/figures/figures.html>
- Quick R website (many helpful “how to” pages)
<http://www.statmethods.net/>
- A U. of Wisconsin Stats professor’s site (B. Yandell)
<http://www.stat.wisc.edu/~yandell/software/>
- Book on Regression and Anova on the R site:
<http://cran.r-project.org/doc/contrib/Faraway-PRA.pdf>