Course Syllabus: MSDS 7333 Quantifying the World, Fall 2016

Course Designer:	Assistant Professor Darren Homrighausen Department of Statistical Science Southern Methodist University, Dallas, Texas E-mail: dhomrighausen@smu.edu
Course Instructors:	Alan Elliott, Daniel Engels, Eric Larson, and Monnie McGee (And Me)
Required Text:	Nolan, D., and Temple Lang, D. (2015), *Data Science in R: A Case Studies Approach to Computational Reasoning and Problem Solving*. Boca Raton, FL: CRC Press (NTL). ISBN: 978-1-4822-3481-7
Suggested Text:	McKinney, W. (2014), *Python for Data Analysis*. San Francisco, CA: O'Reilly Publishers, available in PDF download (McKinley). ISBN: 978-1-449-31979-3
Prerequisites:	Last semester of MSDS program

Learning Objectives: The student will

- Be able to form a testable hypothesis from an unstructured problem
- Create an analysis plan to test aforementioned hypothesis
- Use code in R, SAS, and Python to perform various advanced analyses
- Learn about some advanced statistical and computational methods to analyze data
- Communicate the findings of a research project in a clear, concise, and scientific manner

Course Overview

Quantifying the World (QTW) is built in two-week chunks of varying subject matter. The first week of each chunk consists of 70 to 90 minutes of theoretical and methodological background that will be necessary to solve the case study for the second week. Topics covered include multiple imputation, branching processes, parallel processing, management of very large data sets, time series, machine learning, and predication of location for indoor positioning systems.

Course Expectations

The Nolan and Temple Lang text (NTL) contains example R code for 12 different case studies. We present 5 of them and the statistical/computational methods to fully understand them in this course. However, due to the ever-changing nature of R, this code doesn't always work as stated in the text. The same is true of the Python code in the McKinley text. You are expected to work out these issues as much as possible.

Grading: Case Studies (50%), Participation (30%), Asynchronous Material (20%)

Case Studies (50%): Every other week in class will be spent working through a case study using R, SAS, or Python. These case studies will be written up and turned in for a grade. Case study write-ups are to be "technical report ready," which basically means that the write-up is suitable for public viewing.

Participation (30%): Attending and contributing to the synchronous sessions.

Asynchronous Material (20%): Watching asynchronous material and answering questions in BLTs as required.

Course Coverage

Unit	Lecture Topic/Case Study	Reading	Software	Presenter
1	Multiple Imputation	Notes on imputation	SAS	Elliott
2	Replacing Missing Data in a Real Study	Case study coding and write-up	SAS	Elliott
3	Time Series in Python	Chapters 5, 10, and 11 of McKinley	Python	McGee
4	Financial Applications of Time Series	Case study coding and write-up	Python	McGee
5	Modeling Signal Attenuation, Cross-Validation, Densities and Distributions	Chapter 1 of NTL and associated readings	R	Engels
6	Predicting Location via Indoor Positioning Systems	Case study coding and write-up	R	Engels
7	Piecewise Linear Regression, local regression (smoothing), Longitudinal Data	Chapter 2 of text of NTL and associated readings	R	McGee
8	Modeling Runners' Times in the Cherry Blossom Race	Case study coding and write-up	R	McGee
9	Naïve Bayes, Recursive Trees, CART, perhaps other methods	Chapter 3 of NTL and associated readings	R	McGee
10	Using Statistics to Identify Spam	Case study coding and write-up	R	McGee
11	Branching Processes and Monte Carlo Simulation	Chapter 7 of NTL and associated readings	R	McGee
12	Simulation Study of a Branching Process	Case study coding and write-up	R	McGee
13	Parallel Computing, Split-Apply-Combine, SQL	Chapter 5 of NTL and associated readings	Python/R	Larson
14	Analyzing Airline Flight Delays: A 12-GB Data Set	Case study coding and write-up	Python/R	Larson
15	Final Project: Rewrite one of the case studies that you have done previously using a different computer language.			

Rescheduling Assignments: Life happens. Instructors schedule major due dates on the same day. Should you need to an extension on an assignment, please let me know at least 24 hours prior to the due date. The notice should be given via e-mail. We will discuss the best course of action given your circumstances. **Note: Issues with technology are NOT an excuse for late work.** You have been warned to start your assignments early enough so that you can resolve such issues before they affect your ability to turn in the work.

SMU Honor Code: Students are expected to abide by the SMU Honor Code, which can be found online at http://www.smu.edu/StudentAffairs/StudentLife/StudentHandbook/HonorCode. The Honor Code prohibits academic sabotage, cheating, fabrication, facilitating academic dishonesty, and plagiarism. Definitions of all of these items can be found on the Honor Code website. The penalty for a first offense is a zero for the assignment. A second offense will result in a failing grade for the course and possible dismissal from the program.

Submission guidelines for assignments

- Your name must be at the top of the first page and on each successive page.
- Submit case studies as a formal written paper. The case study should have an abstract, an introduction, a literature review, a methods section, a results section, and a future work/discussion/conclusion section. Code should be included in an appendix to the document. Spelling and grammar count!
- Use an easy-to-read variable-width font with a minimum of 11-point font. (I like Arial, Helvetica, and Geneva fonts—this document is in Helvetica 11 point.)
- Relevant code and output must be included in-line at the appropriate point using Courier New (or other fixed width) font, in 10-point size. **Inclusion of irrelevant code or output, even in an appendix, will be penalized**. All software output must be given in the text or as a table created in Word (or the software you are using).
- Any graphics must be electronically cut and pasted in-line at the appropriate point of the write-up. You can use Word to resize the graphics appropriately. **Screen shots from SAS, R, or Python tabular output are not allowed in the document text or in the appendix.** All tables and figures should have descriptive titles and captions. In short, the reader should be able to understand the content of the figure or table without reading the associated text.
- Any mathematical notation must be provided with appropriate use of subscripts, superscripts, and symbols. Use MS Equation or another equation editor if you submit your work in Word.

Best Practices for Success in MSDS 7333

Attendance. Take responsibility for your commitment. Attendance means not only being there for synchronous sessions but also participating in asynchronous work.

Citizenship. You need to be actively engaged to succeed in this class. Talking on cell phones, texting, "Facebooking," tweeting, or leisure web browsing are prohibited in class. I consider these to be a disruption (not to mention rude).

Integrity. A lot of the graded work occurs outside of class, so I expect honesty and integrity in what you submit for evaluation. Evidence of academic dishonesty will minimally result in zeros for all involved parties, and perhaps University-level disciplinary action. Don't risk your academic career.

Humility. Don’t get lost! Ask questions in class. If something isn't clear to you, it probably isn't clear to others either. Questions may arise because I haven’t made a connection clear or have inadvertently left out an important point. Your question gives me a chance to explain more clearly. Don't be proud or shy.

Organization. Don’t procrastinate! This is a technology-driven course. Count on your computer failing or your wireless connection breaking the night before a due date. Start early and give yourself a chance to succeed.

Deadlines. You will generally have a week to complete an assignment. Due dates and times will be clearly indicated. Late submissions will be penalized, but it is much better to turn in work late than not at all (or to turn in incomplete/sloppy work). Work turned in after solutions have been posted to the course website will receive no credit.

Getting help. If questions arise while doing assignments/exams, do your best to resolve these questions before the assignment is due, first by taking time to seek answers yourself, and then via e-mail to your instructor or other students. **I encourage you and expect you to seek help.**

Collaboration. I encourage the formation of study groups and collaboration with your fellow students in tackling the assignments. Working together in groups on homework is permitted, even encouraged. **However, every student should write up and complete his or her homework independently. Students who chose to turn in exactly the same work will share the grade assigned.** Talking about problems with other people does help in learning, but just copying the solutions from one another doesn't help!

Looks do matter! All assignments must be NEATLY executed and organized. You risk a zero on any assignment submitted in a sloppy manner. See submission guidelines for more detail.

University Policies

Grading Policy: Graduate students must receive a C or better in a course in order to pass the course. If a student must retake a course, then the second grade and the first grade are averaged for the purposes of the overall GPA. Failure to maintain a GPA of 3.0 or better will result in dismissal from the program.

Incompletes will be given only in the case of extraordinary circumstances that prevent you from finishing the semester. You must have completed at least 50% of the course with a passing grade to be eligible for an incomplete.

Religious Observance: Religiously observant students wishing to be absent on holidays that require missing class should notify the live session instructor via e-mail, and should discuss with the instructor, in advance, acceptable ways of making up any work missed because of the absence. (See University Policy No. 1.9.)

Excused Absences for University Extracurricular Activities: Students participating in an officially sanctioned, scheduled University extracurricular activity will be given the opportunity to make up class assignments or other graded assignments missed as a result of their participation. It is the responsibility of the student to make arrangements with the instructor prior to any missed scheduled examination or other missed assignment for making up the work (University Undergraduate Catalogue).