IST687 Applied Data Science
School of Information Studies
Syracuse University

1. Course Description

The course introduces students to applied examples of data collection, processing, transformation, management, and analysis to provide students with hands-on introduction to data science experience. Students will explore key concepts related to data science, including applied statistics, information visualization, text mining and machine learning. "R", the open source statistical analysis and visualization system, will be used throughout the course. R is reckoned by many to be the most popular choice among data analysts worldwide; having knowledge and skill with using it is considered a valuable and marketable job skill for most data scientists.

1.1. Course Structure

During the synchronous section, you will have a chance to practice and apply you knowledge via an assignment that is due before the next synchronous session. Beyond the homework, there will also be a large project that is part of this course. This project will allow you to apply what you have learnt within the class to a real-world data problem. You will work within a team for this project, and be provided a real-world dataset. Your team's task is to understand the domain and the data available to determine how to best provide insight and wisdom from all the data that might be available.

1.2. Learning Objectives

At the end of the course, students are expected to understand:

- Essential concepts and characteristics of data
- Scripting/code development for data management using R and R-Studio
- Principles and practices in data screening, cleaning, and linking
- Communication of results to decision makers

At the end of the course, students are expected to be able to:

- Identify a problem and the data needed for addressing the problem
- Perform basic computational scripting using R and other optional tools
- Transform data through processing, linking, aggregation, summarization, and searching
- Organize and manage data at various stages of a project lifecycle
- Determine appropriate techniques for analyzing data

1.3. What does it take to succeed in the course?

- An interest and passion in data science - in the corporate, academic, or government sector
- Curiosity about business, science, education, health or another substantive area
- Essential computer skills particularly around spreadsheets
- Close familiarity with algebra, geometry, and trigonometry

- Basic understanding of simple descriptive statistics
- Motivation to learn and achieve a high degree of professional preparation

2. Textbooks:

Introduction to Data Science (2017), by Jeffrey S. Saltz & Jeffrey M. Stanton is available from Sage -
http://us.sagepub.com/en-us/nam/an-introduction-to-data-science/book256486

The book is also available in in paperback and electronic version on Amazon.com
https://amazon.com/Introduction-Data-Science-Jeffrey-Saltz/dp/150637753X/

I will provide additional and supplemental readings in the LMS as electronic documents for downloading and printing. Students are expected to read the assigned materials for discussions and coursework.

3. Contributions to Grade

The work for this class will involve the following:

- **HW** (30%) are designed for you to practice the necessary skills in carrying out data processing, analysis, and management tasks.
- **Class Participation** (12%) includes attendance and participation in-class.
- **Mid-term quiz** (30%) is designed to evaluate your mastery of concepts, methods, and tools in data analysis and management.
- **Final project** (28%): For the final project you work on a dataset provided, transform the data as needed, and provide a written analysis with visualizations (group of 4-7 students). Students will be assigned into a group. The grade is comprised of:
    - 25% final submission
    - 3% for the project updates (1% for each of the 3 project updates submitted)

4. Grading Policy

- Each assigned work will be graded on the scale as specified for the component, which will be summed at the end of the semester.
- Grade levels follow the scales below:

Highest	Lowest	Letter	Expectation
100.00 %	93.00 %	A	Your work is outstanding
92.99 %	90.00 %	A-	
89.99 %	87.00 %	B+	Your work is about what would be expected of a serious student
86.99 %	83.00 %	B	
82.99 %	80.00 %	B-	
79.99 %	77.00 %	C+	Your work falls below what is expected but is adequate
76.99 %	73.00 %	C	
72.99 %	70.00 %	C-	
69.99%	0.00 %	F	Your work is out of the picture

- It is unethical to allow some students additional opportunities, such as extra credit assignments, without allowing the same options to all students.
- Students who wish to dispute a grade may resubmit the assignment for regrading with a one-page statement of explanation of why the paper should be regraded. If the student resubmits, the assignment will be regraded, which means the grade may go up, down, or stay the same. Except for extraordinary circumstances, no appeal for an individual assignment or project will be considered later than two weeks after the assignment was graded.

5. Schedule - This is a is subject to change

Date	Readings**	Topics+	Topic during Lab*	Homework*
Week 0				Install R
Week 1	Ch 1: About Data Ch 2: Data Problems Ch 3: Getting Started with R	What is Data Science & R Overview	Basic R coding (vectors, conditionals)	HW 1: Working with Vectors
Week 2	Ch 4: Follow the Data Ch 5: Rows and Columns Ch 6: Data Munging	Using R to manipulate data.	Data Frames & sorting	HW 2: Manipulating data frames
Week 3	Ch 7: Onward with R-Studio Ch 8: What's my Function? Ch 9: Beer,Farms,& Peas	Descriptive Statistics & Functions	Descriptive Stats & Functions	HW 3: Distributions & writing functions
Week 4	Ch 10: Sample in a Jar	Inferential statistics	Sampling & Decisions	Project Update 1 HW 4: Sampling & Decisions
Week 5	Ch 11: Accessing Data	Connecting with external data sources	External Data: JSON	HW 5: Getting Data
Week 6	Ch 12: Pictures vs numbers	Introduction to visualization	Data Viz	Project Update 2 HW 6: Visualizations
Week 7	Ch 13: Map Mash-Up	Working with map data	Maps	HW 7: Working with Maps
Week 8	Ch 16: Line Up, Please	Linear modeling	Quiz	HW 8: Linear modeling
Week 9	Ch 17: Hi Ho, Hi Ho-Data Mining We Go Ch 18: What's your vector, Victor?	Association Rule Mining and Support Vector Machines	Using exploratory analysis and arules	Project Update 3 HW 9: Support vector machines
Week 10	Ch 14: Word Perfect Ch 15: Happy Words?	Text Mining	Text Mining	HW 10: Text Mining
Week 11			Final Project	

+Topics discussed in Video Lectures

*Assignments must be submitted by midnight on the night prior to the start of the next synchronous class.

**Readings should be complete prior to video lecture

General Requirement for Assignment Submissions

Assignments must be professionally prepared with computer applications. Unless otherwise stated, assignments must be submitted electronically to the LMS. No hand-written assignments will be accepted. An assignment must be in one document when it is submitted to the LMS. If you have additional supporting materials that are in physical forms or hard copies (e.g., business forms or some images), you must scan them into JPG or TIFF format and embed them into your MS Word or PDF document. If you use MS Word to prepare your assignments, use Times New Roman style with 12-pt font, 1.5 line spacing, and 1 inch margin around.

For each assignment, be sure to include the a title page with the following information:

- Course number (i.e. IST 687)
- Name(s) of the author(s) or the team members
- Name of the assignment (e.g. "Homework 2")
- Date the assignment is due and date in which it is submitted
- Number pages in the body of the assignment (not necessary on the first page)

You must submit all assignments to the LMS on the deadline specified for each assignment.

HW assignments are due prior to the start of the next synchronous class. Late HW assignments will receive a 0 (no credit). If your final project is late, **I will deduct 10% of the original grade for the first day of lateness plus 15% for every subsequent day.** You will not receive full credit for topics/assignments presented in an unprofessional manner. Professionalism includes the proper use of grammar, punctuation, and limiting spelling mistakes. Professionalism also means adhering to given instructions. Failure to adhere to the assignment instructions will result in a reduction of the grade. If English is not your first language set up an appointment with the writing program so they can help you improve your writing.

Respect and Disruption

It is expected that we all treat people's contributions and differences of opinion with respect. There are certain actions that can be disruptive not only to your own learning experience but to everybody else's as well. Examples include talking to neighbors during class, arriving late, cell phone ringing during class, text messaging, falling asleep, reading newspapers or magazines, lack of civility and respect in comments made, etc. Your repeated disruption will reduce your final grade. In extreme cases, you can be asked to leave the class and even excluded from the course.

6. University and School Policies

Academic Integrity

The academic community of Syracuse University and of the School of Information Studies requires the highest standards of professional ethics and personal integrity from all members of the community. Violations of these standards are violations of a mutual obligation characterized

by trust, honesty, and personal honor. As a community, we commit ourselves to standards of academic conduct, impose sanctions against those who violate these standards, and keep appropriate records of violations. The academic integrity statement can be found at: http://supolicies.syr.edu/ethics/acad_integrity.htm.

Faith-based Observances

Syracuse University recognizes the diverse faith traditions represented among its campus community and supports the rights of faculty, staff, and students to observe according to these. This link http://supolicies.syr.edu/studs/religious_observance.htm provides a description of SU's religious observance policy. Under this policy, students are provided an opportunity to make up examination, study, or work requirements that may be missed due to religious observance *provided they notify the university and their instructors before the end of the second week of classes.* Students will have access to an online notification system for this purpose on MySlice during the first two weeks of the semester. The make up of an activity affected by a religious observance will be scheduled to be completed within a week of the missed deadline in agreement with the instructor.

Disabilities

If you believe that you need accommodations for a disability, please contact the Office of Disability Services (ODS), http://disabilityservices.syr.edu, located in Room 309 of 804 University Avenue, or call (315) 443-4498 for an appointment to discuss your needs and the process for requesting accommodations. ODS is responsible for coordinating disability-related accommodations and will issue students with documented disabilities Accommodation Authorization Letters, as appropriate. Since accommodations may require early planning and generally are not provided retroactively, please contact ODS as soon as possible.

Ownership of Student Work

This course may use course participation and documents created by students for educational purposes. In compliance with the Federal Family Educational Rights and Privacy Act, works in all media produced by students as part of their course participation at Syracuse University may be used for educational purposes, provided that the course syllabus makes clear that such use may occur. It is understood that registration for and continued enrollment in a course where such use of student works is announced constitutes permission by the student. After such a course has been completed, any further use of student works will meet one of the following conditions: (1) the work will be rendered anonymous through the removal of all personal identification of the work's creator/originator(s); or (2) the creator/originator(s)' written permission will be secured. As generally accepted practice, honors theses, graduate theses, graduate research projects,

dissertations, or other exit projects submitted in partial fulfillment of degree requirements are placed in the library, University Archives, or academic departments for public reference.