IST687 Applied Data Science
School of Information Studies
Syracuse University
DRAFT – Fall 2020

Class Sessions	Larger Class: Virtual - students need to keep Thursdays, 9:30 AM – 10:50 AM free for class meetings	Smaller Labs: Location / online or offline Vary by instructor/section
Instructors:	Jeff Saltz	Email: jsaltz@syr.edu
	Erik Anderson	Email: esande01@syr.edu
	Chris Dunham	Email: cndunham@syr.edu

Office Hours: Upon request, using Zoom.

Important notes:
- You must own a functioning Windows or Mac laptop and bring it to every lab session.
- Your laptop must have R and R-Studio installed (or you need to have an account on https://rstudio.cloud/)
- You must obtain the textbook in digital or print form by the first class, please have it for all lab sessions (the textbook contains example code and explanations that will be helpful for the lab).
- Email to the professor should have the subject line "IST687: XYZ" where XYZ summarizes your question.
- If emailing about homework, please include a complete copy of your R code file as an attachment to the message.

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

The class is a combination of asynchronous material (e.g., pre-recorded discussions) and live sessions. Specifically, each week will have pre-recorded video material - lectures, roundtables and coding demonstrations. The assigned chapters from the textbook should be read prior to reviewing the pre-recorded video material.

Each week, you will have a chance to practice and apply your knowledge via a homework assignment and in in-class lab assignment. In addition, each week, you will also submit lecture

notes and questions that have come up during the week. All lab sections will be live (check blackboard for the details for your lab session). The weekly lab synchronous sessions will review key concepts, let you practice R coding and will also review any student questions that have arisen during the week.

Beyond the homework and lab assignment, 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, where your 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.

Finally, there will also be an exam that will include questions about the pre-recorded material and will also include R-coding questions.

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:

The primary textbook for this class is *Introduction to Data Science (2017)*, by Jeffrey S. Saltz and Jeffrey M. Stanton (ISBN: 978-1506377537). The book is available in in paperback and electronic version on Amazon.com:

https://amazon.com/Introduction-Data-Science-Jeffrey-Saltz/dp/150637753X/

As needed, additional and supplemental readings will be provided 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:

- **Homework (22%)** - The 11 homework assignments (each worth 3 points) are designed for you to practice the necessary skills in carrying out data processing, analysis, and management tasks. The code you create for the homework assignments will be instrumental in helping you to complete the exam successfully.

- **Lecture Notes (12%)** - While you review the 12 asynchronous Module material, you must take written notes on what you learn. At the end of each module you will submit a copy of your written notes on Blackboard. You will receive one point for each set of notes you submit, as long as the notes fit the rubric presented below.

- **Lab Assignment (11%)** - While you are in lab, you will work on an in-class assignment. At the end of each lab session you will submit a copy of your work (R code) on Blackboard. You will receive one point for your submission, as long as you have given a reasonable effort (i.e., it is OK if the code was not perfect – the lab section is where you are trying to learn the R coding concepts).

- **Attendance Class Participation (5%) –** Did you actively participate in your lab session? This is a course composed of content that you will consume weekly in an asynchronous fashion and weekly face-to-face live class sessions. Your active and constructive involvement with the asynchronous content and live class sessions will help to ensure that you receive the most benefit from this class. In other words, active participation is important. Like the professional environment, it is expected that you will be in class every week and that you will consistently appear on time. You are expected to ask questions and seek clarification when something in unclear to you. Your participation grade (**5%** of final grade) is based on the following elements over the 11 weeks:
    1. Attendance in the live sessions.
    2. Active participation in the live sessions by contributing content, asking questions, and taking part in the group breakouts. Participation can be in the chat box or verbally.
    3. This includes the homework presentation within the lab section.

    Your instructor recognizes that your participation may vary week to week and that sometimes life circumstances can interfere. Therefore, the grade is based on your participation over the semester.

- **Exam (25%)** – The exam takes place during your lab session. It will be designed to take 90 minutes to complete. The exam will be open-book and open-notes, but you may not discuss the exam with other students. You will need your laptop and R/R-Studio to complete the exam.

- **Final project (25%) –** Predictive Analytics effort. Will be done during the last half of the course. More information will be made available during the course.

4. Code Template for Homework

All homework assignments are ".R" files (source code for R) that you will prepare in RStudio. Assignments must be submitted electronically to Blackboard. No hand-written assignments will be accepted. All R code must be commented according to the standards described by the instructor. All homework assignments should begin with the following code template:

```
####################################################
# IST687, Standard Homework Heading
#
# Student name:
# Homework number:
# Date due:
#
# Attribution statement: (choose only one)
  # 1. I did this homework by myself, with help from the book and the professor
  # 2. I did this homework with help from the book and the professor and these Internet sources:
  # 3. I did this homework with help of <Name of person> but did not cut and paste any code

# Run these three functions to get a clean test of homework code
cat('\014') # Clear the console
rm(list=ls()) # Clear user objects from the environment

# Set working directory - Change to the folder containing your homework data files
setwd("~/Desktop/IST687/Homework")
```

5. The Large Class

The large class will not meet on Thursdays. However, you do need to continue to block that time on your calendar. That time will be used, for example, for meeting (online) with your lab instructor outside of the lab time (with your team). In addition, part of that time will be for your lab instructor's office hours (this will ensure everyone will be available during office hours, if needed).

6. Lecture Notes

Lecture notes must be submitted for each module. Besides serving as evidence that you reviewed the pre-recorded material, the notes that you take will help you complete the exam successfully. Here is a rubric documenting the key components of an acceptable submission of lecture notes:

1. Each module of pre-recorded material matches with the chapter or chapters that were assigned from the book. Your notes must contain at least one summary statement for each assigned book chapter. Summary statements for a chapter should mention and describe each major section of the chapter.
2. Each module of pre-recorded material contains several different videos: Your notes must have at least one summary statement for each video.
3. Some of the videos, particularly the ones labeled as "lecture," contain a slide with one or more written questions. Your notes must contain an answer to these questions (it is OK if your answer is not correct, as long as you explain your answer).
4. Each week of pre-recorded material contains at least one video pertaining to coding with R/R-Studio. Your notes should contain at least one R command and an explanation of what it does.

5. Any questions you have, with respect to the content, should be put into the discussion on that module (within blackboard). The answers to these questions will be provided in lab (or via blackboard).

The criteria described above are minimal – that is they are the minimum requirements needed to earn the credit for the module notes submission. Your learning process and your performance on the exams will be better served by taking more extensive notes.

7. Lab Sessions

During most lab sessions, you will work on a lab coding assignment, typically with one other person. No specific pre-work is required for the lab coding assignment – ***but you need to come to class prepared (i.e., will likely need to use your text book, R code examples and lecture notes)***.

In addition, several times during the semester, as part of a team, you will provide a short presentation describing the solution for the assigned homework. For example, during week 3 class on Sept 10th, a team will present the homework solution for week 2. The team will need to meet with the lab instructor prior to that lab session (either during the large lecture time or another agreed upon time). For each presentation, another team will be responsible to evaluate the presentation of the team and a third team will be accountable to ask at least two questions relating to the presentation. These presentations will be a key aspect of your participation grade.

8. Schedule - This is subject to change

Week	Date of Lab**	Readings*	Key Focus of Lab	Homework***
1	Aug 27	Ch 1: About Data Ch 2: Data Problems Ch 3: Getting Started with R	R Overview / What is Data Science	HW 1
2	Sep 3	Ch 4: Follow the Data Ch 5: Rows and Columns Ch 6: Data Munging	Using R to manipulate data (data frames & sorting)	HW 2
3	Sep 10	Ch 7: Onward with R-Studio Ch 8: What's my Function? Ch 9: Beer, Farms & Peas	Descriptive Statistics & Functions	HW 3
4	Sep 17	Ch 10: Sample in a Jar	Inferential statistics (sampling & decisions)	HW 4
5	Sep 24	Ch 11: Accessing Data	Connecting with external data sources	HW 5
6	Oct 1	Ch 12: Pictures vs numbers	Introduction to visualization	HW 6
7	Oct 8	Ch 13: Map Mash-Up	Working with map data	HW 7
8	Oct 15	Ch 16: Line Up, Please	Linear modeling	HW 8
9	Oct 22	Ch 17: Hi Ho, Hi Ho-Data Mining We Go	Association Rule Mining	HW 9
	Oct 25		Exam (includes Linear modeling, but not Association Rule Mining)	
10	Oct 29	Ch 18: What's your vector, Victor?	Support Vector Machines	HW 10
11	Nov 5	Ch 14: Word Perfect Ch 15: Happy Words?	Text Mining	HW 11
12	Nov 12		Leading teams / Ethics + Final Project	
13	Nov 19		Final Project	
	Dec 3			Final Project

*Readings and watching the lecture/discussion should be complete prior to writing lecture notes.
**The lecture notes should be completed prior to your lab session, lab work submitted after the lab.
***Then after your lab session, do the homework (which is due 3 days after the lab session)
Points will be deducted if assignments are late.

9. **Grading Policy**
   1. Each assigned work will be graded on the scale as specified for the component, which will be summed at the end of the semester.
   2. 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

   3. It is unethical to allow some students additional opportunities, such as extra credit assignments, without allowing the same options to all students.

   4. 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.

10. **SPECIAL NOTICES RELATED TO THE COVID-19 PANDEMIC**
During the COVID-19 pandemic, norms for the conduct of classes and for student classroom behavior are necessarily different than in the past.

Stay Safe Pledge
Syracuse University's Stay Safe Pledge reflects the high value that we, as a university community, place on the well-being of our community members. This pledge defines norms for behavior that will promote community health and wellbeing. Classroom expectations include the following: wearing a mask that covers the nose and mouth at all times, maintaining a distance of six feet from others, and staying away from class if you feel unwell. Students who do not follow these norms will not be allowed to continue in face-to-face classes; repeated violations will be treated as violations of the Code of Student Conduct and may result in disciplinary action.

Food and Drink in the Classroom
Eating and drinking require the lowering of the face mask, creating a potentially dangerous situation. For this reason, students are not allowed to eat or drink in class during the COVID-19 pandemic.

Online Etiquette
Students participating remotely in hybrid class sessions are participating on an equal basis as the students who are in the classroom. It is reasonable to expect online students to have their

webcam on and to call on them as you would call on students who are in the room. Online students should dress and behave as they would in a face-to-face class.

Attendance
The university's Stay Safe Pledge requires students to stay away from in-person classes when they feel unwell. Therefore, as needed, students can join remotely for their lab section, even if it is "their turn" to physically attend the lab session.

11. 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.

Syracuse University's Academic Integrity Policy reflects the high value that we, as a university community, place on honesty in academic work. The policy defines our expectations for academic honesty and holds students accountable for the integrity of all work they submit. Students should understand that it is their responsibility to learn about course-specific expectations, as well as about university-wide academic integrity expectations. The policy governs appropriate citation and use of sources, the integrity of work submitted in exams and assignments, and the veracity of signatures on attendance sheets and other verification of participation in class activities. The policy also prohibits students from submitting the same work in more than one class without receiving written authorization in advance from both instructors. Under the policy, students found in violation are subject to grade sanctions determined by the course instructor and non-grade sanctions determined by the School or College where the course is offered as described in the Violation and Sanction Classification Rubric. Syracuse University students are required to read an online summary of the University's academic integrity expectations and provide an electronic signature agreeing to abide by them twice a year during pre-term check- in on MySlice.

Course evaluations
At the end of the term, the iSchool will ask you to share course feedback through EvaluationKIT [https://coursefeedback.syr.edu]. Log in to EvaluationKIT using your NetID and password. Please take the time to share your feedback about this course and your experience in it; all ratings and comments are completely anonymous. The iSchool carefully reviews your feedback. Our instructors use this feedback to fine tune course delivery and instruction; our professors of record use this feedback to fine tune course content and assignments. All feedback is factored into iSchool decisions about course, program and instructor development.

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.