

COURSE SYLLABUS - DRAFT
IST 687: Intro to Applied Data Science

1

Term: Fall 2021

Large Lecture Delivery Mode: In-person

Large Lecture Day: Thursday

Large Lecture Time: 9:30am – 10:50am

Location: TBD

Lecture Instructors:

Jeff Saltz, jsaltz@syr.edu

Lab Sessions: Check Blackboard

Office Hours: upon request

Lab Instructors (tentative):

Erik Anderson, esande01@syr.edu

Chris Dunham, cdunham@syr.edu

Yang Yang, yyang87@syr.edu

Jesse L Cases, jlcases@syr.edu

Ayse Dalgali, ayocal@syr.edu

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.

Course Description:

The course introduces students to fundamentals about data and the standards, technologies, and methods for organizing, managing, curating, preserving, and using data. It discusses broader issues relating to data management, ethics, quality control and publication of data. Applied examples of data collection, processing, transformation, management, and analysis as well as a hands-on introduction to the emerging field of data science are provided. 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.

Additional Course Description:

COURSE SYLLABUS - DRAFT
IST 687: Intro to Applied Data Science

2

The course will consist of one larger class plus one smaller lab section each week. During the larger class lectures, we will explore key concepts and examine the use of those concepts within R. Prior to attending lecture, it is expected that you've completed the reading and the preparatory exercise that is due the evening before the lecture. The lecture and lab will be easier, faster, more fun, and more useful if you get as much as you can out of the lecture. Preparation and participation are key.

The weekly lab sessions will review key concepts, let you practice R coding and will also review any student questions that have arisen during the week. During the lab and the homework, you will have a chance to practice and apply your knowledge.

There will also an **exam** that will include questions about the lecture material as well as R-coding questions.

Note that there will also be a final **project** (group) which you will complete during the semester. 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.

Audience:

Graduate students

Credits:

3 credit hours

Learning Objectives:

After taking this course, the students will be expected to understand:

- Essential concepts and characteristics of data
- The purpose of scripting for data management using R and R-Studio
- Principles and practices in data screening, cleaning, linking, and visualizations
- The importance of clear communication of results to decision-makers
- The key ethical challenges associated with applications of data science in a variety of contexts

After taking this course, the students will 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

Texts / Supplies – Required:

The primary textbook for this class is *Data Science for Business With R (2021)*, by Jeffrey S. Saltz and Jeffrey M. Stanton (ISBN: 978-1544370453). The book is available in paperback and electronic version at the University bookstore and on Amazon.com:

<https://www.amazon.com/Data-Science-Business-Jeffrey-Saltz/dp/1544370458>

Texts / Supplies– Additional:

Additional and supplemental readings may be posted in Blackboard as electronic documents for downloading and printing.

Course Requirements and Expectations:

Students are expected to read the assigned materials prior to each class and lab session.

Students are expected to take notes on what they read and also on the information from the lecture and the lab. Good notes are key to success in this course.

Lab Sessions

During most lab sessions, you will work on a lab coding assignment, perhaps with one other person. No specific pre-work is required for the lab coding assignment – ***but you need to come to class prepared.***

Grading:

The work for this class will involve the following:

- **Homework (33%)**—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 (Quiz-based) Challenges (11%)**— You will have an opportunity to participate in class challenges (quiz) in the form of brief (4 to 6 multiple choice questions). You need to do these quizzes after each lecture (must be completed by 11:59AM on that Thursday).
- **Lab Assignments (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).
- **Class Participation (5%)**—Did you actively participate in your lab session.
- **Mid-term Exam (20%)**—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 (20%)**—Predictive Analytics effort. Will be done during the last half of the course. More information will be made available during the course.

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:

Grading Tables

Letter Grade	Percentage Range	Expectation
A	100.00%-93.00%	Your work is outstanding, always complete, and submitted on time.
A-	92.99%-90.00%	
B+	89.99%-87.00%	Your work is what we expect from a serious, professional student, but may contain occasional errors, omissions, or inconsistencies.
B	86.99%-83.00%	
B-	82.99%-80.00%	
C+	79.99%-77.00%	Your work falls below what we expect but is adequate for completing the course.
C	76.99%-73.00%	
C-	72.99%-70.00%	
F	59.99%-0.00%	Your work is insufficient to pass the course.

Course Specific Policies:

1. It is unethical to allow some students additional opportunities, such as extra credit assignments, without allowing the same options to all students.
2. 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.

How to Succeed in This Course:

- Cultivate an interest and passion in data science—in the corporate, academic, or government sector
- Show curiosity about business, science, education, health or another substantive area
- Have essential computer skills particularly around spreadsheets
- Be familiar with algebra, geometry, and trigonometry
- Have basic understanding of simple descriptive statistics
- Be motivated to learn and achieve a high degree of professional preparation
- Honor and uphold the values of the iSchool (see below)

SYRACUSE UNIVERSITY STUDENT POLICIES & SERVICES

See the Syllabi Appendix for information on iSchool values as well as Syracuse University Policies.

COURSE SYLLABUS - DRAFT
IST 687: Intro to Applied Data Science

6

Course Schedule:

subject to change

Week/ Date of Lecture	Lecture Topic	Required Readings	Focus of Lab	Assignments*
0/ Aug 26	Class Introduction			
1/ Sep 2	Data Science Overview	Ch 1: Begin at the Beginning With R	Install R/RStudio, Using R	HW 1 Assigned
2/ Sep 9	Using R to manipulate data	Ch 2: Rows and Columns Ch 3: Data Munging	Working with Dataframes	HW 2 Assigned
3/ Sep 16	Descriptive Statistics & Functions	Ch 4: What's my Function? Ch 5: Beer, Farms & Peas	Using Descriptive Statistics & writing functions	HW 3 Assigned
4/ Sep 23	Sampling & Inferential Statistics	Ch 6: Sample in a Jar	Sampling	HW 4 Assigned
5/ Sep 30	Connecting with External Data Sources	Ch 7: Accessing Data (storage wars)	Using JSON	HW 5 Assigned
6/ Oct 7	Introduction to visualization	Ch 8: Pictures vs numbers	Using GGPLOT2	HW 6 Assigned
7/ Oct 14	Working with map data	Ch 9: Map Mash-Up	GGPLOT2 with maps	HW 7 Assigned
8/ Oct 21	Linear Modeling (LM)	Ch 10: Lining Up Our Models	Working with LM	HW 8 Assigned
X/ Oct 28	<i>No Lecture – Study for Exam</i>		Exam in Lab	
09/ Nov 4	Supervised Learning	Ch 11: What's your vector, Victor?	Support Vector Machines and Trees	HW 9 Assigned
10/ Nov 11	Unsupervised Learning	Ch 12: Hi Ho, Hi Ho-Data Mining We Go	Association Rules Mining	HW 10 Assigned
11/ Nov 18	Text Mining	Ch 13: Word perfect (Text Mining)	Word clouds and Sentiment Analysis	HW 11 Assigned
XX / Nov 25	Thanksgiving Break			
12/ Dec 2	Ethical Challenges & Interactive Apps	Ch 14: Shiny Web Apps	Final Project	
13/ Dec 9	Deep Learning	Ch 15: Time for a Deep Dive	Final Project	

*HW assignments are **due 6 days after they are assigned** (i.e., by the Wed following when they were assigned. **You must submit all assignments to Blackboard prior to the deadline.** They will be accepted until 11:59 PM of the due date.