COURSE SYLLABUS SPRING 2024

IST 387: Introduction to Applied Data Science

Lecture Instructor: Chris Dunham; cndunham@syr.edu

Lecture Meetings: Thursdays, 11:00AM-12:20PM, Maxwell Hall Auditorium

Office Hours: Request appointment by email unless otherwise posted to Blackboard

Lab Meeting	Lab Instructor	Email
Tu 11:00AM-12:20PM, Hinds Hall 011	Preeti Jagadev	pjagadev@syr.edu
Tu 6:30PM-7:50PM, Hinds Hall 011		
Tu 12:30PM-1:50PM, Hinds Hall 011	Brittany Johnson	bjohns06@syr.edu
Tu 5:00PM-6:20PM, Hinds Hall 011	Jason Reilly	jsreilly@syr.edu

Important Notes:

- You must own a functioning laptop with **R** and **RStudio** installed or create an account on ***posit.cloud*** (formerly *rstudio.cloud*).
- You must bring your laptop to every class meeting, both lecture and lab.
- You must obtain the textbook in digital or print form by the first class and are encouraged to bring it to each class (the textbook contains example code and explanations that will be helpful for the lab portion of class).
- Emails addressed to the professor should have the subject line "**IST 387: 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.

COURSE SYLLABUS SPRING 2024

IST 387: Introduction to Applied Data Science

Additional Course Description:

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. 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 be 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: Undergraduate 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

COURSE SYLLABUS SPRING 2024

IST 387: Introduction to Applied Data Science

Course Schedule (subject to change)**:**

Week/ Date of Lecture	**Lecture Topic**	**Required Readings**	**Focus of Lab**	**Assignments***
Before first class	-	-	-	Install RStudio, Do lab 0
01/ Jan 18	Data Science Overview	Ch 1: Begin at the Beginning With R	Using R	HW 1 - Assigned
02/ Jan 25	Using R to Manipulate Data	Ch 2: Rows & Columns Ch 3: Data Munging	Working with data frames	HW 2 - Assigned
03/ Feb 1	Descriptive Statistics & Functions	Ch 4: What's my Function? Ch 5: Beer, Farms & Peas	Descriptive statistics & writing functions	HW 3 - Assigned
04/ Feb 8	Sampling & Inferential Statistics	Ch 6: Sample in a Jar	Sampling	HW 4 - Assigned
05/ Feb 15	Connecting with External Data Sources	Ch 7: Accessing Data (storage wars)	Using JSON	HW 5 - Assigned
06/ Feb 22	Introduction to visualization	Ch 8: Pictures vs numbers	Using ggplot2	HW 6 - Assigned
07/ Feb 29	Working with map data	Ch 9: Map Mash-Up	Using ggplot2 with maps	HW 7 - Assigned
08/ Mar 7	Linear Modeling (LM)	Ch 10: Lining Up Our Models	Working with LM	HW 8 - Assigned
XX/ Mar 14	***No Class – Spring Break***	-	-	-
09/ Mar 21	Supervised Learning	Ch 11: What's your vector, Victor?	Support vector machines and trees	HW 9 - Assigned
10/ Mar 28	Unsupervised Learning & Ethical Challenges	Ch 12: Hi Ho, Hi Ho-Data Mining We Go	Association rules mining	HW 10 – Assigned
XX/ Apr 4	***No Lecture – Study for Exam***	-	**Exam in lab**	-
11/ Apr 11	Interactive Apps	Ch 14: Shiny Web Apps	Shiny	HW 11 – Assigned
12/ Apr 18	Text Mining & ChatGPT	Ch 13: Word perfect	Word clouds and sentiment analysis	HW 12 – Assigned
13/ Apr 25	***No Lecture – Group Project Time***		Final project	-

*HW assignments are due **6 days after they are assigned** (i.e., each Wednesday night 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.