STAT 3055 INTRODUCTION TO R FOR DATA SCIENCE (3 credit hours)

Elmira College

SPRING 2025

Required Text:

Grolemund and Wickham (2023): ***R for Data Science*** (2nd ed.). O'Reilly.
Grolemund (2014): ***Hands-On Programming with R***. O'Reilly.
Supplemental readings might be included to illustrate or expand on textbook readings.

Pre-requisites: MATH 1535 Calculus with Analytic Geometry I; STAT 2114 Introductory Data Science

Course Description

This course introduces students to the fundamental concepts in data science, covering the entire data science workflow, various aspects of statistical and machine learning techniques. It explores the R programming language and R packages for data manipulation, visualization, and modeling. Through hands-on laboratory sessions, students will engage in practical exercises, turning raw data into meaningful insights, knowledge, and understanding, and effectively communicating analytical results using R, RStudio and R Markdown.

Course Objectives and Goals

- Understand the fundamental concepts and principles of data science.
- Develop proficiency in R and data science tools for data manipulation, analysis, and reporting.
- Apply basic statistical techniques and machine learning algorithms.
- Create effective data visualizations for communication.
- Work collaboratively on data science projects.

Evaluation of Performance

Your grade will be based upon your performance on exams, assignments, and participation.

Assignments	15%
Labs	20%
Midterm Exam	15%
Final Project	25%
Final Exam	25%
Total	100%

Grades will be assigned as follows:

A	93% and above	B-	80 - 82%	D+	67 - 69%
A-	90 - 92%	C+	77 - 79%	D	63 - 66%

B+	87 - 89%	C	73 - 76%	D-	60 - 62%
B	83 - 86%	C-	70 - 72%	F	59% or below

Withdrawal Policy: Please see Elmira College Bulletin for information on this policy.

Academic Honesty: Please read the section on Academic Honesty in the **Code of Conduct**. Briefly, academic dishonesty includes: cheating, fabrication, facilitating academic dishonesty, and plagiarism. Ask if you have any questions on whether something constitutes as academic dishonesty. All work must be original and new. Past assignments from current or other courses will not be accepted. Academic dishonesty will not be tolerated. It will result in zero on the assignment, and a report will be filed with the school. Continued practice will result in failure of the class. Institutional penalties may also apply with repeated acts of academic honesty.

Student Responsibility:

- It is your responsibility to keep track of assignments and due dates.
- You should ask questions concerning assignments and lectures, if you need any clarifications.
- If you are struggling in class, have concerns, and/or unsure about expectations, please stop by during office hours or make an appointment for another time.

Tentative Schedule of Topics

Topic	Materials	Tasks & Evaluations
Introduction to R, RStudio, R Markdown	Chapter 1	
Data Visualization	Chapter 2	Assignment 1
Data Types and Representation	Chapter 3	
Data Frames and Data Manipulation	Chapter 4	
Data Transformation with Dplyr	Chapter 5	Assignment 2
Data Summarization and Pipes	Chapter 6	
Date Import and Workflows	Chapter 7	Lab 1
Tidy Data, Relational Data	Chapter 8	
Graphics for Communication with ggplot2	Chapter 9	Midterm Exam
R Markdown Formats	Chapter 10	
Web Scraping with R (rvest Package)	Chapter 11	Lab 2
A Field Guide to Base R	Chapter 12	
Machine Learning in R	Chapter 13	Assignment 3
Supervised Learning: Classification and Regression	Chapter 14	
Unsupervised Learning: Clustering and Dimensionality Reduction	Chapter 15	Lab 3
Functions and Conditional Execution	Chapter 16	
Classification Techniques in R	Chapter 17	Lab 4
Dates and Times	Chapter 18	Final Project
Iteration and Functional Programming	Chapter 19	Final Exam