

# STAT/CS 2870 Online: Basics of Data Science

## Summer 2025

**Instructor:** Jacob Martin

**E-mail:** [Jacob.W.Martin@uvm.edu](mailto:Jacob.W.Martin@uvm.edu)

**Office Hours:** By appointment

### Course Description:

Basic techniques of data harvesting and cleaning; association rules, classification, clustering; analyze, manipulate, visualize data using programming languages. Basic principles of probability and statistical modeling/inference to make meaning out of large datasets. Prerequisite: STAT 1110/1410/2430

### Detailed Description:

Basic data science techniques, from import to cleaning to visualizing and modeling, using the R language. Machine learning methods include regression and classification algorithms. Programming methods include loops, indexing, and user-defined functions.

This course is an introduction to the field of Data Science for students with **no experience in computer coding**.

### Learning Outcomes:

- Write code in the R programming language, using the R Studio integrated development environment.
- Use R base package to index objects, use loops and conditioning, write functions, and create model objects and lists.
- Use R base package and the dplyr package to input data, read in files, and shape data.
- Use R base package and the ggplot2 package to create a variety of data visualizations.
- Using R Markdown, organize and document code to enhance effectiveness and readability.
- Interpret data visualizations; identify characteristics of effective and aesthetic visualizations.
- Understand the basics of modeling, and be able to employ basic modeling techniques, such as linear regression, in appropriate situations
- Understand the basics, and the strengths and weaknesses of common algorithmic approaches, including classification algorithms using nearest neighbors and decision trees; regression trees and model trees; and clustering with k means
- Recognize ethical considerations in the use of algorithms and 'big data.'
- Communicate effectively both results and procedures.
- Identify ways that data can be misrepresented and create data summaries that represent data accurately and effectively.

### Textbook:

Data Visualization: A Practical Introduction, by Kieran Healy. Princeton University Press, 2019.

ISBN: 978-0-691-18162-2 (You may also choose to use an "Incomplete Draft" of the book, available online at [socviz.co](http://socviz.co))

Machine Learning with R, by Brett Lantz, 2<sup>nd</sup> edition. (required, but 2<sup>nd</sup> edition available free online through UVM library)

R for Data Science, by Garrett Golemund and Hadley Wickham. Available free online:

<https://r4ds.had.co.nz/introduction.html> (extra resource for reference, we'll use a little)

### Technology:

Access to a computer is required for the course. We will be using Brightspace for the course homepage and where you can find the material for the class. While there are no meeting times, we will use Microsoft Teams (MS Teams) for office hours. MS Teams is free for anyone enrolled at UVM. Lastly, we will be using R and RStudio to program, which is always free, regardless if you are enrolled at any university!

### **Programming:**

This course includes a basic introduction to programming in the computer language [R](#); using the work environment, [RStudio](#). No programming experience is required. R is **free** and has many useful add-on packages that we'll be using in the course.

### **Homework Assignments:**

There will be two homework assignments due per week, due on Wednesday and Friday. The homework assignments and any necessary files are located in their corresponding modules in Brightspace

Each assignment will be submitted as a pdf in Gradescope using the Gradescope link in Brightspace.

The homework assignments will have different point totals, and your homework grade will add the total points earned to form your homework grade. This means not all homework assignments are worth an equal amount towards your homework grade.

Example: If you earned a 25/27 on HW1 and 27/30 on HW2, your homework grade would be  $(25+27)/(27+30) = 52/57 = 91\%$

We will be dropping 5% from the total points possible for the homework grade.

You can submit your solutions as many times as you like before the due date and the last one submitted will be graded.

*No late assignments will be accepted, and you'll want to be sure you've submitted the correct file before the due date. If what you've submitted isn't the correct file or file type, it won't be graded!*

### **Final Project:**

There will be a small project at the end of the semester for you to illustrate what you've learned using a data set from a list provided in Brightspace. You can find more details about the project in the Project section.

### **Email:**

All email should be through your UVM email account. I check my email Monday through Friday between 8 AM and 5 PM. If you email about an assignment on Friday after 5, it is unlikely you will get a response before the following Monday, after the due date and late submission time.

### **Late Work, Grade Disputes, and Extra Credit:**

The course will only last 6 weeks, so we will be moving at a fast pace. Solutions to the homework assignments will be posted immediately after the due date. Because of these two factors, no late work will be accepted.

Any disputes over grades need to be made within 1 week after grades have been posted.

Final letter grades will be determined as seen below. No individual adjustments will be made and there is no "extra credit" work.

**Grade Distribution:**

Assignment	Total Weight
Homework	65%
Project	35%

**Letter Grade Distribution:**

Letter	+		-
A	97 - 100	92.5 - 96.9	89.5 - 92.4
B	86.5 - 89.4	82.5 - 86.4	79.5 - 82.4
C	76.5 - 79.4	72.5 - 76.4	69.5 - 72.4
D	66.5 - 69.4	62.5 - 66.4	59.5 - 62.4
F	0-59		

**Expectations and UVM Policies:**

Students are expected to act in accordance to the rules outlined in the [University's Policy Statements](#).

- **Religious Holidays:** If you need to miss class to observe a religious holiday, please submit the dates of your absence in writing by the end of the second full week of the semester.
- [Academic Integrity Policy](#)
- [Code of Student Rights and Responsibilities](#)
- [FERPA Rights Disclosure](#)
- **Diversity** at [UVM](#) and in the [College of Engineering and Mathematical Sciences](#)

**STUDENT LEARNING ACCOMMODATIONS:**

In keeping with University policy, any student with a documented disability interested in utilizing accommodations should contact SAS, the office of Student Accessibility Services on campus. SAS works with students and faculty in an interactive process to explore reasonable and appropriate accommodations, which are communicated to faculty in an accommodation letter. All students are strongly encouraged to meet with their faculty to discuss the accommodations they plan to use in each course. A student's accommodation letter lists those accommodations that will not be implemented until the student meets with their faculty to create a plan. Contact SAS: A170 Living/Learning Center: [802-656-7753](tel:802-656-7753) (phone link).