

DATA 180-01: Introduction to Data Science

Fall 2024, Tome Hall 121

TF 3:00 p.m. – 4:15 p.m.

Instructor: Dr. Zachary Kessler

Email: kesslerz@dickinson.edu

Office: Althouse G19

Campus Phone: 717-254-8082

Office Hours: Monday and Thursday 12:30-2:30 p.m. or by Appointment

QRA: Malena Malka Goldman

Email: malkagom@dickinson.edu

Office Hours:

Location:

Course Description

The ability to work with and derive information from ever-increasing amounts of data will be one of the important stories of the 21st century. New analytical techniques coupled with rapidly advancing computational power continues to change way data is collected, organized, analyzed, and understood. A facility with data science techniques allows the student to bring this exciting new toolkit to bear helping to mine information from almost every area of human interest. DATA 180 provides an introduction to the core ideas of data science. Topics include data visualization, data wrangling, statistical measures of center, spread, and position, and supervised and unsupervised statistical/machine learning. Upon successful completion of the course a student will be able to:

- Organize, manipulate, and transform data using R,
- Use Github and RMarkdown to create reproducible reports and maintain a repository for version control,
- Analyze and interpret data using visualization techniques and statistical summaries,
- Employ supervised and unsupervised machine learning techniques for predictive modeling,
- Identify internal structure in data organize, manipulate, and transform data in a statistical programming environment,
- Comprehend and create basic numerical and/or logical arguments.

We will make extensive use of the R and R-Studio to generate graphical and numerical

representations of data, and apply basic machine learning techniques while we interpret the results. R is a fun and useful computational tool as well as an immediate resume builder

Required Materials:

MATH 180: Introduction to Data Science Course Packet by Jeff Forrester, available at the Dickinson College bookstore (required)

Introduction to Statistical Learning by G. James, D. Witten, T. Hastie, R. Tibshirani (recommended)

Spatial Data Science with Applications in R by E. Pebesma and R. Bivand (optional)

Reasoning with Data by Jeff Stanton (optional)

Access to a computer to install and use R

Here is the link to our course [GitHub](#).

At various points during the semester, I will assign short news stories, blog posts, or academic articles. Links to these will be provided to you via official channels on Moodle, Teams, or GitHub at no charge.

Grades:

Assignments 20%

Exam 1 25%

Exam 2 25%

Final Exam 30%

Assignments: Homework assignments will be posted on course Github page as an R-Markdown

file template on which you will insert your solutions. Due dates will be provided for each assignment. You will turn in your assignments as an R-Markdown file via a pull request from your private GitHub.com repository which is a clone of the class master repository. (You will need to set up a GitHub account if you do not already have one.) You will be sent an invitation

link for each assignment. After accepting the assignment, your private repo where you will push your files will automatically be created. Prior to pushing your submission files to

your repository, make sure to hit Knit on R-Studio, and include the .Rmd file in your commit.

Make sure your code executes with no issues. You will receive a 20% penalty if any part of your code cannot get executed because of errors. Email submissions will not be accepted. Late assignments will not be graded.

Exams: There will be three exams in this class. These will cover the topics and subjects discussed during the particular, relevant unit. These will contain some mixture of multiple-choice, short answer, and coding questions. You will be allowed a single, one-sided, handwritten note sheet in exams.

Class Outline:

Below is a tentative schedule and reading list for our course:

Week 1: Introductions and Getting Started in R
Course Notes, 3-6, Textbook, 1-9

Week 2: Data, Variables, Functions, Oh my...
Course Notes, 6-16, Textbook, 15-28 (Stop at 2.1.5)

Week 3: Visualization: Bars, Pies, and Histograms
Course Notes, 12-29

Week 4: Measures of Centrality and Relationships
Course Notes, 30-58

Week 5: ggplot Visualizations

Week 6: EXAM 1

Week 7: Data Wrangling and Dplyr
Course Notes, 63-76

Week 8: Unsupervised Learning: Hierarchical Learning
Course Notes, 95-129, Textbook, 519-528

Week 9: Kmeans Clustering
Course Notes, 133-148, Textbook, 515-518

Week 10: Mixed Variables Clustering
Course Notes, 149-163

Week 11: Text Analysis

Week 12: EXAM 2

Week 13: Linear Regression
Course Notes, 179-192, Textbook, 59-102

Week 14: Trees-Based Methods
Textbook, 327-352

Week 15: Final Review

FINAL EXAM, MAY 13th 9:00 a.m. – 12:00 p.m.

Grading Scale:

Letter Grade	Score
A	93+
A-	90-92.9
B+	87-89.9
B	83-86.9
B-	80-82.9
C+	77-79.9
C	73-76.9
C-	70-72.9
D+	67-69.9
D	63-66.9
D-	60-62.9
F	<60

Quantitative Reason Associate for the Course:

Dickinson College provides additional support for students taking courses with quantitative content across the curriculum through the Quantitative Reasoning (QR) Center. Our Quantitative Reasoning Associate (QRA) for this semester is Malena Malka Goldman. As a fellow student, she has experience in this course and providing additional assistance

throughout the class. She will provide support through additional office hours and review sessions prior to exams.

Late Work Policy:

Assignments can be submitted one class after their expected due date (i.e. if a homework assignment is due Tuesday, you may submit at class the following Friday) at the penalty a letter grade. Exams cannot be submitted late unless extenuating circumstances require a student to take the exam at a later date. Two options are available to students in the above situation. First, you may simply take the exam at a later date to be arranged with me. Second, you may move the weight of the exam to the final, effectively doubling its worth.

Laptop and Phone Policy:

Laptops will be allowed to be used during class for notetaking/coding purposes and visiting relevant topical websites I specify. I reserve the right to move to paper-only notetaking and laptops only for coding use if students are found to be repeatedly using laptops beyond these specific use cases.

Phones will not be allowed under any circumstances in class. If you need to make a call, please step outside.

Gen AI Policy:

The utilization of AI tools such as LLM's like OpenAI's ChatGPT, Google's Gemini, Meta's Llama 3, or Anthropic's Claude will be permitted on a highly selective basis. You may only use these tools on assignments where I explicitly permit them. These tools are not allowed for exams.

Absence Policy:

Attendance of this course is mandatory and repeated, unexcused absences will harm the final grade. I will track and count the number of absences. If you need to miss class, notify me once you are aware of the conflict. Traditionally applicable excuses will be accepted (i.e. medical/family emergency, sudden, unforeseen circumstances, etc.). This policy is in effect because participation will be used to reward students who consistently contribute to the positive learning environment of the course.

Dickinson Honor Code:

Students are expected to follow Dickinson's Honor Code in every aspect of the course. No cheating or plagiarism will be tolerated for any assignments. Any individual caught using such methods will be reported to the necessary parties and be subject to Dickinson's Student Conduct process.

Statement on Accommodations:

Dickinson values diverse types of learners and is committed to ensuring that each student is afforded equitable access to participate in all learning experiences. If you have (or think you may have) a learning difference or a disability – including a mental health, medical, or physical condition– that would hinder your access to learning or demonstrating knowledge in this class, please contact Access and Disability Services (ADS). They will confidentially explain the accommodation request process and the type of documentation needed to determine your eligibility for reasonable accommodations. To learn more about available supports, go to www.dickinson.edu/ADS, email access@dickinson.edu, call (717)245-1734, or go to the ADS office in Room 005 of Old West, Lower Level (aka "the OWLL").

Physical Location of Course and Professor's Office

If you've already been granted accommodations at Dickinson, please follow the guidance at www.dickinson.edu/AccessPlan for disclosing the accommodations for which you are eligible and scheduling a meeting with me as soon as possible so that we can discuss your accommodations and finalize your Access Plan. If you will be using any test-taking accommodations in this class, be sure to enter all test dates into your Access Plan in advance of our meeting. ADS will be happy to provide any assistance you may need.

My office is located on the floor of garden level of Althouse, which has an elevator, located to the right of the main entrance. If you require the use of an elevator to access the garden level, please let me know, and be sure that ADS (access@dickinson.edu) knows as well. If there is ever a malfunction with the elevator, I will gladly arrange to meet you at an alternative location, either in the building, nearby, or by video conference.

Statement on Inclusivity

This class will cultivate an inclusive environment supporting a broad variety of experiences and backgrounds. These differences are to be acknowledged, celebrated, and provide insight on any topics discussed.

Student Privacy

During the course of this class, no student's personal information will be disclosed to anyone. For more information on your rights to privacy in this class under FERPA, please visit this link: https://www.dickinson.edu/info/20088/registrars_office/264/ferpa.

SOAR Academic Resources:

SOAR stands for Strategies, Organization, & Achievement Resources, and students can find a wealth of strategic academic success tools (like weekly planners, semester calendars, and much more) at www.dickinson.edu/SOAR or by going to Old West's Lower Level (aka "the OWLL"). Online or in person, you'll find useful resources related to organization, study skills, memory strategies, note-taking, test-taking, etc. If you'd like to attend a SOAR workshop or request one-on-one assistance with developing a strategy for a manageable and academically successful semester, email SOAR@dickinson.edu, or go to the OWLL to SOAR!