MATH/COMP/DATA 180: Introduction to Data Science

Fall 2021, Section 2

(Last updated December 6, 2021; syllabus is subject to change)

Instructor: Professor Eren Bilen
Office: Rector North 1309
Email: bilene@dickinson.edu
Phone: 717-254-8162
Office Hours: Monday 7:00-8:00pm (Zoom)
Tuesday 3:00-4:00pm (in-person)
Thursday 2:00-3:00pm (in-person)
or by appointment
QRA: Lan Pham, phaml@dickinson.edu
Office Hours: Tuesday 3:00-4:00pm (in-person)
Wednesday 7:45-8:45pm (in-person)
Location: Rector North 1311

Class: Tome 121
Tuesday and Thursday
10:30-11:45pm

Class Notes and Other Required Materials

- MATH 180: Introduction to Data Science Course Packet by Professor Jeff Forrester, available at the Dickinson College bookstore (required)
- Access to a computer to install and use R
- Course webpage: Moodle

Course Goals

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. Math 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 learning. Upon successful completion of the course a student will be able to:

- Organize, manipulate, and transform data in a statistical programming environment;

- 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 statistical programming environment to generate graphical and numerical representations of data and complete statistical learning techniques while we interpret the results. R is a fun and useful computational tool as well as an immediate resume builder!

Course Policies

Attendance Policy: This course will be taught in person in Tome 121. Students are expected to attend all in-class meetings, which occur on Tuesdays and Thursdays from 10:30-11:45pm ET. While I will not take formal attendance, it is important for you to attend in class meetings and take notes. If you will be unable to attend a class meeting for any health-related issues or other emergencies, please contact me beforehand so that arrangements can be made.

Use of Laptops, Tablets, and Phones: Laptops, tablets, and phones are not permitted during this course as they can be highly distractive in the classroom. To access R, you are free to use any of the desktop computers in the room. We will usually work in pairs. Phones should be put away in silent mode.

Grading: Your course grade is based on two closed-book midterms, a final project in place of a final exam (more information to come), and weekly homework.

Midterm 1 (20%):	October 14
Midterm 2 (20%):	December 2
Final project (20%):	by December 15
Homework (40%):	Due dates TBA

While I will not be giving "extra credit", I will drop your lowest homework. I expect there to be 12 total assignments (depending on course pacing). Occasionally, an assignment may be weighted to count as two assignments (because of the complexity or length), this will be clearly indicated when it is assigned.

The following scale will be used to determine your final grade:

Score	Letter	GPA	Score	Letter	GPA
$93 \geq x$	A	4.0	$73 \leq x < 77$	C	2.0
$90 \leq x < 93$	A-	3.7	$70 \leq x < 73$	C-	1.7
$87 \leq x < 90$	B+	3.3	$67 \leq x < 70$	D+	1.3
$83 \leq x < 87$	B	3.0	$63 \leq x < 67$	D	1.0
$80 \leq x < 83$	B-	2.7	$60 \leq x < 63$	D-	0.7
$77 \leq x < 80$	C+	2.3	$x < 60$	F	0.0

Make-up Exams: There will be no make-up exams unless a student must be away from campus on university business or due to an emergency. The student must provide documentation.

Homework: Homework will be assigned in Moodle on a weekly basis. Due dates will be provided for all assignments. Solutions will be made available in Moodle after the deadline. You may discuss homework problems with others, but your solutions must be written up individually. You are expected to justify your answers in complete sentences except for questions that are clearly pure computations. Homework assignments will require the use of R; you will want to obtain access to a machine with R installed during the first week of classes; R is installed in the various labs in Tome Hall.

Homework assignments must be turned in person at the beginning of class as a stapled physical copy. All homework assignments must conform to the `attached guidelines`. Assignments not conforming to these standards may be penalized. Late assignments will not be accepted without a signed medical excuse. If you are unable to submit an assignment because of a personal or health emergency that arises during the week, you must contact me prior to the submission time to discuss arrangements for completing the assignment. Your work will be scored and returned in class.

Final Project: The course will include a final data analysis project in lieu of a final exam that will be due Wednesday, December 15 at 5:00 pm EDT. More information will be posted later in the semester.

Getting Help

Office Hours: I will be holding office hours in-office and in Zoom–please see Moodle for the Zoom link that will be used throughout the semester. My office hours are Monday 7:00 – 8:00pm (Zoom), Tuesday 3:00–4:00 (in-person), and Thursday 2:00 – 3:00pm (in-person). I am also available by appointment.

Quantitative Reasoning Associate: This semester, we are fortunate to have a Quantitative Reasoning Associate (QRA) working with us. A QRA is a fellow student who completed the course in the past and will be helping us as a course facilitator and mentor. This semester, the QRA for our course is Lan Pham. Lan has office hours on Tuesday 3:00-4:00pm and Wednesday 7:45-8:45pm, both in person, at Rector 1311.

If you need to attend office hours on Zoom, please contact Lan ahead of time to arrange this. Lan will host exam review sessions, which will be announced closer to each exam.

Quantitative Reasoning Center

Dickinson College provides additional support for students taking courses with quantitative content across the curriculum through the Quantitative Reasoning (QR) Center. For the fall 2021 semester, the QR Center will offer tutoring for DATA 180, in addition to general quantitative support. You are strongly encouraged to make an appointment with them. Click here to access the QR Center webpage.

Please visit dickinson.mywconline.com to make an appointment. Then, access the drop-down menu under "limit to" at the top of the scheduler and select DATA 180. This will restrict the tutor list and schedule to only those tutors approved for this course. When you make your appointment, please also paste or upload your assignment and any work that you have done.

Other Important Information

Referencing the Work of Others: When submitting your work, you must follow common-sense ground rules. External sources may only be used to improve your own understanding of the material. When you write your solutions, you should do it on your own without the direct help of any external sources, and certainly should not write down anything that you do not understand. If you do use external references, please be sure to cite them. Failure to cite references will be treated as academic dishonesty.

Respect for Intellectual Property: It is important that you be aware of and respect the intellectual property rights of others. Unless explicitly stated otherwise, all materials available on the Internet, in libraries, and elsewhere are considered intellectual property and can only be used with the permission of the owner. Specifically, with regards to this class, you should not share any of the course materials, including homework answer keys, with others, even after the completion of the course.

Statement on Disabilities: 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 impairment 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 that Dean and Director Marni Jones will need 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").

If you have 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 test proctoring will be needed from ADS, remember that we will need to complete your Access Plan in time to give them at least one weeks advance notice.

SOAR: Academic Success Support: Students can find a wealth of strategic guidance by going to www.dickinson.edu/SOAR. This website for SOAR (Strategies, Organization, and Achievement Resources) includes apps, tips, and other resources related to time management, study skills, memory strategies, note-taking, test-taking, and more. You will also find information aimed to helps students "SOAR Through Academic Challenges," as well as a schedule of academic success workshops offered through Academic Advising. If you would like to request one-on-one assistance with developing a strategy for a manageable and academically successful semester, email SOAR@dickinson.edu.

Course Outline: Below is a list of topics to be covered in this course. There may be adjustments on the list during the semester depending on progress. Any adjustments will be announced on Moodle.

- Chapter 1: Introduction to Data Science

- Chapter 2: Data and Variables

- Chapter 3: Displaying Distributions

- Chapter 4: Numerical Summaries of Data

- Chapter 5: Data Transformations

- Chapter 6: Introduction to Data Wrangling

- Chapter 7: Unsupervised Learning: Cluster Analysis

- Chapter 8: Visualizing Multidimensional Data

- Chapter 9: Probability

- Chapter 10: Introduction to Supervised Learning

Important Dates for the Fall 2021 Semester

Last Day to Add/Drop or Change to/from Pass/Fail	Friday, September 3
Mid-Term Pause	5 pm, Friday, October 15 thru 8 AM, Wednesday, October 20
Course Request Period for Spring 2022 Semester	Monday, November 1 thru Wednesday, November 3
Thanksgiving Vacation	5PM, Tuesday, November 23 thru 8 AM, Monday, November 29
Last Day to Withdraw from a Course with a "W" grade	Tuesday, November 23
Classes End	Friday, December 10
Reading Period Days	December 11, 12

Homework Guidelines

The importance of homework cannot be overstated as it provides the main vehicle for truly understanding the material presented in this course. Therefore, you must give top priority to working on and understanding homework problems. The instructions below provide some basic guidelines for creating organized, legible, and professional looking homework.

Why Format Matters? There are a number of reasons why I require carefully formatted homework assignments:

- Learning statistics involves learning how to communicate your ideas effectively. As a student, much of this communication will be in the form of homework. How you present your work should enhance the ideas you are trying to communicate.

- Proper formatting allows me to provide meaningful feedback.

- Grading assignments is much more efficient when they are formatted well.

Requirements

1. Homework assignments should be typed or handwritten in pencil. All handwriting should be neat, organized, and legible.

2. Do not cross out errors; erase them.

3. Use 8.5×11 white paper. ("printer-paper", letter-size)

4. There should be no perforated edges. No twisted corners.

5. Leave 1-inch margins on all four sides of the paper and leave white space between problems.

6. Homework assignments must be stapled. There is not a stapler in the classroom and therefore you should staple your homework before coming to class.

7. Avoid scratchwork on assignments: instead, first work out the solutions to homework problems on scratch paper, and then write them up neatly.

8. All R code and output should be in a `fixed-width font`, in which all the characters have the same width. This ensures that all indentation and spacing is retained. Courier is an example of such a font.

9. If you are asked to include R output, all comments related to the output should appear directly next to the output. Ideally, you should copy all R output into Microsoft Word or Latex and then type your comments immediately following the output. Use your judgement for figure sizes.