

Instructor: Jeremiah W. Johnson, Associate Professor, Dept. of Applied Engineering & Sciences

Office Hours: TR 9:10 - 10:00 AM, Room 105

Phone: (603) 641-4127

Email: jeremiah.johnson@unh.edu

Class Hours: M 1:10 PM - 3:00 PM in P361

Discovery Requirements: This course meets the university's Discovery program requirement in the area of Environment, Technology and Society (ETS).

Textbooks: You will have regular assigned readings from the two books listed below. Both books are freely available as e-books through the UNH library. The ISBN numbers below are for the print editions of the books, should you wish to purchase hard copies instead of reading them online. Readings from other sources will be assigned regularly.

1. *Learning Python*, by Lutz, 2013. ISBN: 978-1-4493-5573-9.
2. *Python for Data Analysis*, by McKinney, 2013. ISBN: 978-1-449-31979-3.

Catalog Course Description: An introduction to data science and analytics. Overview of the use of analytics by industry, government, and nongovernmental organizations. Impact of analytics on society, ethical use of analytics. Methods of data generation, data management, data cleaning, and data preparation, with a focus on visual and exploratory analysis. Project-based, with an emphasis on collaborative, experiential learning. Programming and statistical software will be used, but previous experience is not required.

Prerequisite: None.

Grading Policy:

Assignments	Value
Homework	30%
Projects	70%

Course grades will be assigned based on your final total percentage as follows:

A: 93 - 100	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: 59.9 or below			

Note that this differs from the standard UNH grading scale, so your final grade may differ from that shown in myCourses!

Credit Hours: This is a 4-credit course. The University of New Hampshire conforms to the U.S. federal definition of a credit hour. According to this definition, a four credit course is expected to involve a minimum of 180 hours of total student work effort (classroom instruction, reading/viewing materials, making notes, doing research, fulfilling assignments, etc.), estimated for an "average student" in this course. Depending on

your prior experience and education, the exact number may be different for you, and the amount of work effort may differ from week to week.

Homework: Homework will be assigned regularly. Homework should be done immediately following the relevant class meeting. Some time will be set aside each week to discuss and review the homework assignment. Homework assignments are typically assigned readings and programming or data analysis mini-projects designed to get you familiar with a specific task, though you will have an occasional nontechnical writing assignment.

Projects: Projects form the bulk of your grade in this course. There will be three–four projects in this course.

Attendance & Participation: We will only be meeting face-to-face once per week. Attendance on these days is crucial and is expected. As of today, attendance and participation will not be graded, but it may be taken into consideration by the instructor when considering a borderline grade. Your instructor may revise this attendance and participation policy if attendance becomes an issue.

Hybrid Course Content: This is a hybrid course, which means that a significant portion of the course will take place asynchronously (that is, at your own pace) online. The hybrid content will consist of a mix of video lectures, readings, exercises, and projects. It is critical that you keep up with the assigned online content. **Failure to keep up with the online content will make it next to impossible to succeed in this course.**

Late/Make-Up Work: Late/make-up work will be considered on a case-by-case basis at the discretion of the instructor. In case of illness or emergency, contact Dr. Johnson as soon as possible (the day of the assessment in all but the most extreme cases) to discuss options.

Extra Credit: Extra credit may occasionally be available through additional homework exercises and/or projects. These additional opportunities to earn credit will be small in scale. Large, last-minute grade-changing extra-credit assignments will not be offered.

Email: Emailing me at the address above is the best way to reach me, though you can also email through myCourses. I try to respond to email within 24 hours on weekdays. **Email response will be substantially slower on weekends/holidays.** At times, it is necessary for me to make announcements through email. It is the responsibility of the student to regularly check his/her UNH email.

Software/Hardware: Data science is a computational field. We are going to be programming quite a bit, using the [Python programming language](#). You do not need to have previous experience programming; we will start at the beginning.

There are three ways to use Python in this course. The first and **recommended** way to use Python in this course is through the web browser, using a free Google software tool called Colaboratory. This will require you to create a Google account if you don't have one already.

The second way to use Python in this course is to install it (along with some additional tools for data science) on your computer. Python can be installed on any computer with a fully functioned operating system (Windows/Mac/Unix/Linux), but it will not work on a Chromebook. This is not recommended unless you already have some programming experience and are comfortable managing and maintaining a Python distribution on your computer.

Required for OPTIONAL local install only:

The Anaconda Python distribution, using Python 3.XXX. <https://www.anaconda.com/downloads>.

May Be Useful for local install only (Python IDEs/Text editors):

PyCharm. <https://www.jetbrains.com/pycharm/>.

Sublime Text 3. <https://www.sublimetext.com/>.

The third way to access Python is through the university's Loaner Laptops. The software listed above is (or should be) installed on all of the machines available in the Learning Commons and on the University's Loaner Laptops, and of course you can access Google's Colaboratory through the web browser as well. This is recommended as a backup in case you forget to bring your laptop to campus on a class meeting day.

Class Cancellations: If class should have to be canceled by the instructor while normal university classes are in session, students will be notified via email. Such notification will not be made when the university closes due to weather or other unforeseen events.

Academic Honesty Policy: Please read the policy at:

<https://catalog.unh.edu/srrr/student-policies-regulations/academic-integrity/> - University's Academic Honesty policy

Course Objectives: Upon completion of this course, students will be able to:

- Describe the landscape of analytics, including those industries and sectors either currently using analytics or expected to do so in the near future;
- Understand how the role of technology in modern society leads to the generation of data, and the opportunities and changes that this causes for individuals, for society, and for industry;
- Acquire data from a variety of sources;
- Perform basic data management tasks, such as combining datasets from different sources and extracting subsets of datasets;
- Perform basic data cleaning and preparation for modeling, including detecting and addressing outliers, handling missing values, and standardizing data;
- Analyze a dataset using a variety of appropriate visualization techniques; and
- Understand and address the numerous ethical and legal issues that arise when handling data.

Tentative Schedule and List of Topics Covered

Week	Topic
1	Introduction to analytics - a broad overview Getting started with Python and Jupyter notebooks
2	Python Programming - NumPy, basic program control
3	Visualizing the distribution of data Summarizing data numerically
4	Introduction to Pandas for data analysis Exploratory Analysis and heuristics for analyzing data
5	More exploratory analysis: dealing with outliers, duplicates Writing functions in Python
6	Python odds and ends - saving data and models, style, PEP8 Covariance and Correlation
7	Hypothesis testing, A-B tests
8	Power, sample size, multiple comparison errors Problems with hypothesis testing
9	Linear Regression
10	Multivariable Regression Modeling with non-numeric data Improving Models
11	Classification
12	Feature Engineering and the Modeling Workflow
13	Data Ethics
14	Big Data

Important Dates

- Aug. 25: First day of class, short (1 hour) session
- Sept. 1: Labor Day, no class session
- Oct. 13: Mid-semester break, no class session
- Oct. 14 - Oct. 17: Dr. Johnson traveling, no office hours
- Nov. 11: Veteran's Day, no office hours
- Nov. 24: Class session held online-only
- Dec. 8: Last day of class

Alterations to Syllabus: Consider this syllabus a general guide. If changes to the syllabus or schedule need to be made as they course progresses, they will be announced in class.