Introduction to Data Science - Prework

This document outlines the prework required for those enrolled in the Introduction to Data Science course.

Prerequisite

Students should have some experience with Python and have familiarity with basic statistical and linear algebraic concepts such as mean, median, mode, standard deviation, correlation, and the difference between a vector and a matrix.

In Python, it will be helpful to know basic data structures such as lists, tuples, and dictionaries, and what distinguishes them (that is when they should be used).

Prework exceptions

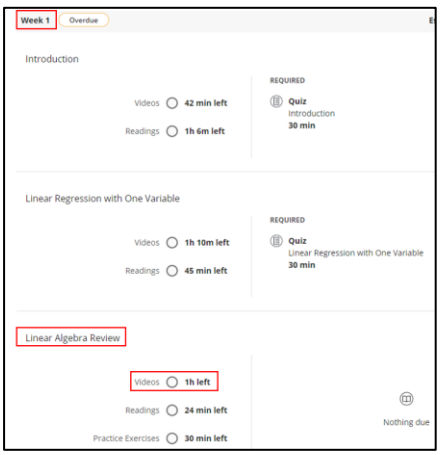
Students should skip the prework if they can accomplish all of the following:

- Write a program in Python that finds the most frequently occurring word in a given sentence.
- Explain the difference between correlation and covariance, and why the difference between the two terms matters.
- Multiply two small matrices together (e.g. 3X2 and 2X4 matrices).

Prework activities

Otherwise, students should complete the following prework (approximately 8 hours) before the first day of class:

- Exercises 1-7, 13, 18-21, 27-35, 38 and 39 of Learn Python the Hard Way.
- Videos 1-6 of Linear Algebra Review from Andrew Ng's Machine Learning course:
    - Enrol for free.
    - Under 'Week 1', refer to 'Linear Algebra Review'.
    - Click on 'Videos'.
    - Complete the 6 videos under this section.



- The exercises in Chapters 2 and 3 of OpenIntro Statistics. This book is free, with a suggested donation. You may donate or set the amount to **zero** to download for free.