CST 383: Intro to Data Science
Dr. Glenn Bruns

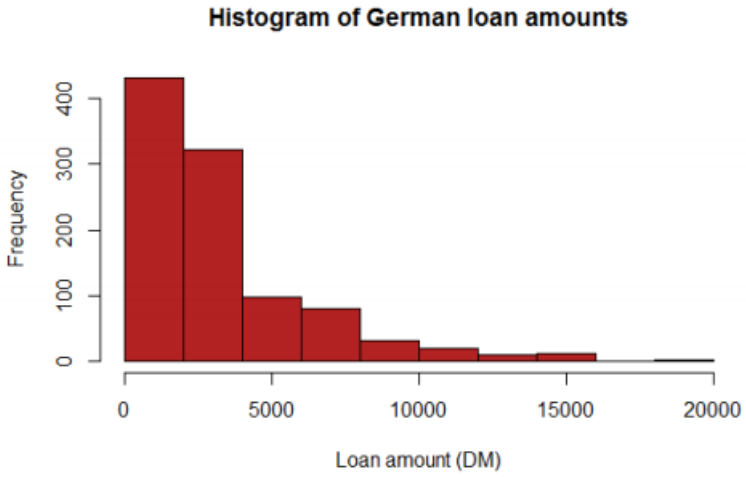
Reading: Data Science Handbook, chapter 2

Purpose. NumPy is used all the time when doing data science with Python. The purpose of this assignment is to help build your initial understanding of NumPy, and to reinforce the key concepts of machine learning we learned this week.

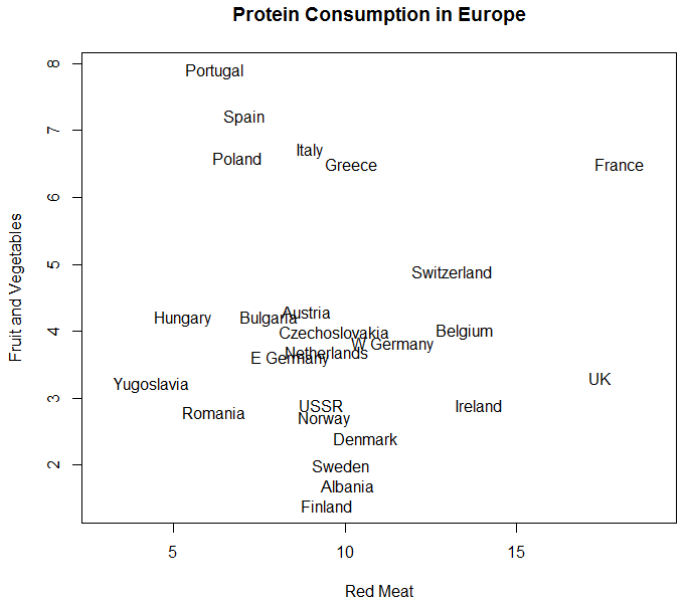
Instructions. Skim Chapter 2 of Python Data Science Handbook (it's available through the O'Reilly Learning Center on the CSUMB dashboard). Get a feeling for what is in the chapter, and read sections in more detail as you like. This chapter will be very helpful for reference.

Answer the following questions by downloading and editing file reading-dsh2.txt (do not change the file name).

1. The plot below is a histogram of German loan amounts. Do you think the amount of most loans in the data is a) more than or b) less than 4000 DM?



2. The plot below is a scatter plot of consumption of two different kinds of food in Europe. Which country has the lowest average consumption of red meat? (capitalization not important)



3. Suppose x is a numeric array of length 10. If you evaluate expression `x < 0`, what is the length of the result?
4. Suppose I want to create an email spam detection system that will predict whether an email is spam or not. Is this a) a regression problem or b) a classification problem?
5. In building my spam detection system, I train the system with some known spam and some known "ham" (email that's not spam). Am I doing a) supervised or b) unsupervised learning?

Submission. Submit your edited reading-dsh2.txt file on iLearn.

Grading. Each problem is worth 10 points.