Let's play with some CSV files. Things to remember / look up from previous code:

importing the csv module

checking to see if a csv file has a header (column headings)

using the next() function to skip a line

If you double click a csv file it's likely to open in numbers (Mac) or Excel (Windows). Remember too that you can open the file in any text editor.

(1) Start by downloading the file GOOG.csv from nexus. Put it in the folder with your lab script.

I downloaded this file of Google stock price and volume from Yahoo finance this morning.

Remember that often data is stored in csv format (comma separated values). That means the data looks something like:

23,24,25,house,zebra

Python is nice, because it presents us a library to help dealing with csv.

Remember to import the csv library, and to open the file for reading as normal, but then also to use the csv.reader method on your open file. This will turn your file into a series of lists that you can loop through, ONE list per line.

You're going to look through this file, one line at a time using a for loop, and try to find some interesting stuff.

(a) Check out the CSV file in excel, numbers or a text editor. Are there column headings? What are they? What order do they appear in? Write this down in comments.

(b) Using python, use a for loop to go through the csv file one line at a time, and print out each line. There are lots of lines, so after a few lines scroll by on the screen, you can interrupt the process by pressing the control key and the c key together (we call this control-c). It should interrupt the program.

You should see the format of each line. What data type is it? NOW look back in excel and look at the column headings. You're interested in several - the date, the price of Google stock at the close of that day, and the volume traded that day. Which columns are they in the csv?

(c) I want you to write code that tells me the following:

- Which day has the highest stock price?

- What's the average stock price?

- What's the average number of google shares traded per day?

- How many days has Google been on the stock market?

HINT: Data type, data types, data types. What data type are the elements in your csv list.

HINT: You'll run into a problem with the first line of the file. Why? What can you write to move to the next line of the CSV file. We did this in class.

HINT: There are MANY ways to solve this. For instance, you could put all the close prices in a list, and then use the max function. BUT - I also want the DAY the max price occurs - so you'll have to find the date that has the max value. OR you can do it all at once - looping over the close values, keeping track of BOTH the biggest value you've seen so far, AND the day on which it happens.

HINT: If you put all the closing values in a list - it's easy to find the average, AND the number of days that Google has been traded.

(2) There is a file on Nexus called Sharks.csv. Use this file and write code that tells me HOW MANY UNPROVOKED shark attacks there have been in the USA since (including) 1990, both in total, and from that the average unprovoked attacks in the USA per year - the data goes up to 2015, so how many per year in the 25 year span?

This file has a slight issue. Different texts are encoded in different ways. Python does it's best to guess the encoding, but this one is tricky. To get it to work properly, we have to tell Python about the specific encoding. To do that, you can use the following line, which adds an encoding argument to the open() function:

data = open(filename, 'r', encoding='ISO-8859-1"')