BIOL5516 Assignment 04

November 23, 2017

0.0.1 BIOL5516 Assignment 04

Created : *Berthin Bitja*

0.0.2 Note :

For each of the following questions, please submit your code and a screenshot of the output. Note that for some questions, there are better (or worse) ways to accomplish a task, and that you should try to find the best solution! Please use comments to indicate which question you are addressing, and to describe what you are doing.

In [2]:
```
# Load the "autoreload" extension
%load_ext autoreload

# always reload modules marked with "%aimport"
%autoreload 1

import os
import sys

# add the 'src' directory as one where we can import modules
src_dir = os.path.join(os.getcwd(), os.pardir, 'src')
sys.path.append(src_dir)

# import my method from the source code
#%aimport preprocess.build_features
#from preprocess.build_features import remove_invalid_data
```

```
The autoreload extension is already loaded. To reload it, use:
  %reload_ext autoreload
```

0.0.3 Ex1

Write a program that reads in the data from "datafile1.txt", which gives the test scores for 10 students in BIOL5516. Calculate the mean test score; make sure that your calculations for finding the mean are contained in a function that you've written.

```
In [38]: def mean(grades):
             average = sum(grades)/len(grades)
             return average

         with open("./_questions/datafile1.txt") as file_object :
             contents = file_object.read()

             # split the data into a list of numbers, remove the last entry which is empty
         students_grade = [ grade for grade in contents.split('\n')][:-1]
             # conver students grade into float integer for calculation
         students_grade = map(float, students_grade)
         s_mean = mean(students_grade)

         print('The mean test score for the 10 students BIO551s is {}'.format(s_mean))
```

```
The mean test score for the 10 students BIO551s is 84.1
```

0.0.4 Ex2

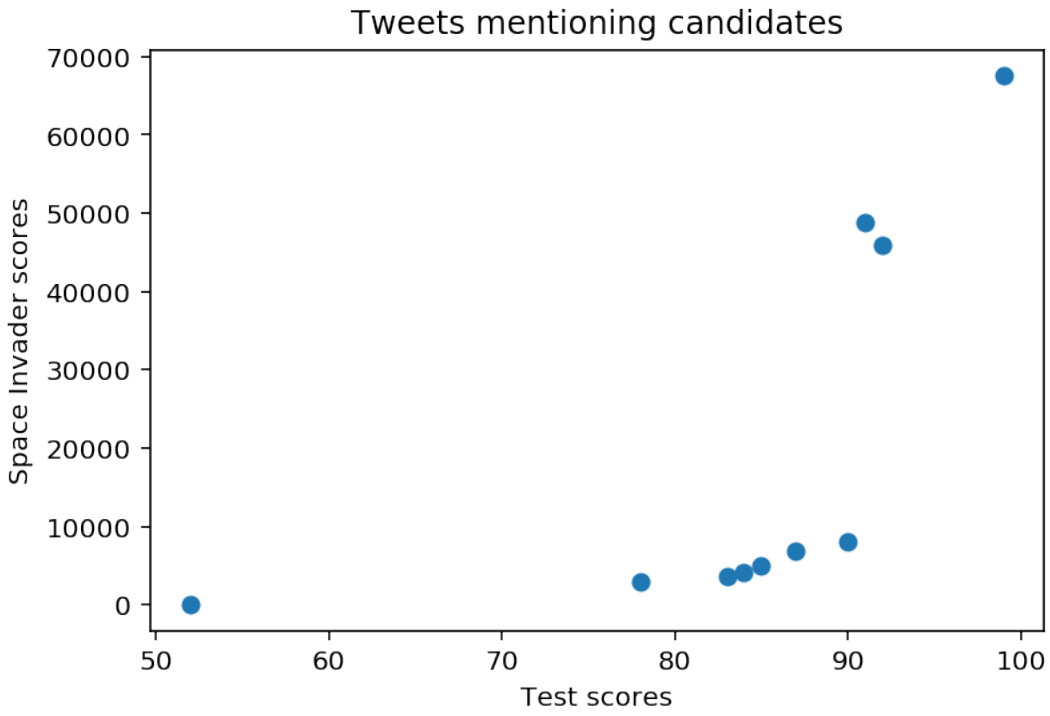
Now consider "datafile2.txt", which gives names, test scores, and Space Invader scores for the 10 students, separated by tabs. Read in the data and calculate the mean test score and the mean Space Invader score. Plot test scores by Space Invader scores. Do you notice anything?

```
In [60]: def mean(grades):
             average = sum(grades)/len(grades)
             return average

         # read the data
         with open("./_questions/datafile2.txt") as f :
             lines = f.readlines()

         #remove new lines at the end
         lines = [ line.strip('\n') for line in lines ]

         #create empty list

         s_names = []
         t_scores = []
         si_scores = []

         for line in lines :
             l = line.split('\t')
             s_names.append(str(l[0]))
             t_scores.append(float(l[1]))
             si_scores.append(float(l[2]))

         t_mean = mean(t_scores)
```

```
        print('The mean test score is {}'.format(t_mean))

        si_mean = mean(si_scores)
        print('The mean Space Invader test score is {}'.format(si_mean))
```

```
The mean test score is 84.1
The mean Space Invader test score is 19300.0
```

```
In [97]: import matplotlib.pyplot as plt
         import numpy as np
         %matplotlib inline
         %config InlineBackend.figure_format = 'retina'
         plt.scatter(t_scores, si_scores)
         plt.title("Test scores by Space Invader scores")
         plt.xlabel("Test scores")
         plt.ylabel("Space Invader scores")
         plt.show();
```



We see that the space score and the test score are positively correlated

0.0.5 Ex3

Now write a program that can report back the data from "datafile2.txt". Your program should read the data into a dictionary, and then prompt the user for a student name. The program should report back the test score and Space Invader score for the student. Remember to check for errors.

```
In [78]: def get_dict():
             #create empty dic
             d = {}
             try:
                 # read the data
                 with open("./_questions/datafile2.txt") as f :
                     lines = f.readlines()
                     #remove new lines at the end
                     lines = [ line.strip('\n') for line in lines ]
                 # read the data into a dictionary
                 for line in lines :
                     l = line.split('\t')
                     print l
                     d[str(l[0])] = [float(l[1]), float(l[2])]
             except FileNotFoundError:
                 return None
             else:
                 return d

         def get_student_score():
             """prompt for the student username"""
             s_name = raw_input("What is your name? ")
             d = get_dict()
             # case 1: test +
             if str(s_name) in d :
                 # report back the test score and Space Invader score for the student
                 print("Hello, {}\n Your score for Space Invader is: {}\n Your score for the st
             # case 2: test -
             else:
                 print("Your name wasn't found in the system")
```

```
In [79]: get_student_score()
```

```
['Trillian', '91', '48900']
['Arthur', '85', '5000']
['Zaphod', '99', '67500']
['Ford', '83', '3600']
['Marvin', '87', '6800']
['Effrafax', '90', '8000']
['Agrajag', '52', '50']
['Fenchurch', '78', '3000']
['Gail', '84', '4150']
['Lunkwill', '92', '46000']
Your name wasn't found in the system
```

```
In [80]: get_student_score()
```

```
['Trillian', '91', '48900']
['Arthur', '85', '5000']
```

```
['Zaphod', '99', '67500']
['Ford', '83', '3600']
['Marvin', '87', '6800']
['Effrafax', '90', '8000']
['Agrajag', '52', '50']
['Fenchurch', '78', '3000']
['Gail', '84', '4150']
['Lunkwill', '92', '46000']
Hello, Arthur
 Your score for Space Invader is: 85.0
 Your score for the student test is: 5000.0
```

0.0.6 Ex4

Write a program that finds the names of all of the files and directories within your current directory, and then reports back the length of each file or directory name.

```
In [96]: import os

         def get_cwd_files(path):
             """return the names of the files and directory in the current directories
             """
             print("my current directory: {}".format(path))
             for file in os.listdir(path):
                 size = os.path.getsize(file)
                 print("{}| size: {}".format(file,size))
         work = os.getcwd()
         get_cwd_files(work)
```

```
my current directory: /Users/bbuildman/Documents/Developer/GitHub/001-BIF5607
_questions| size: 288
BIO5516 Assignment 03.py| size: 1687
BIO5607 - Chap 06.ipynb| size: 16130
BIOL5516 Assignment 04.ipynb| size: 32887
ass_3Q2.py| size: 1185
BIO5516 Assignment 03.html| size: 254113
.DS_Store| size: 6148
ass_3Q3.py| size: 726
BIOL5516 Assignment 02.ipynb| size: 10056
Untitled.ipynb| size: 72
BIO5516 Assignment 03.ipynb| size: 2551
_courses| size: 480
dna.py| size: 456
BIOL5516 Assignment 01.ipynb| size: 9230
BIO5516 Ass 03.py| size: 518
README.md| size: 618
_classes| size: 128
```

```
module.pyc| size: 785
.ipynb_checkpoints| size: 256
module.py| size: 533
.git| size: 512
ass_3Q4.py| size: 568
.idea| size: 256
```