

```
In [1]: import numpy as np
import pandas as pd
import scipy as sp
```

```
In [2]: %matplotlib inline
import matplotlib.pyplot as plt
plt.style.use('ggplot')
```

```
In [3]: %%file hw_data.csv
id,sex,weight,height
1,M,190,77
2,F,120,70
3,F,110,68
4,M,150,72
5,O,120,66
6,M,120,60
7,F,140,70
```

Writing hw\_data.csv

## Python

### 1. Finish creating the following function that takes a list and returns the average value.

Add each element in the list to `total` and return `total`

**DO NOT use a library function nor `sum()`**

```
In [4]: def average(my_list):
        total = 0
        for item in my_list:
            total = total + item
        total = total/len(my_list)
        return total
```

```
average([1,2,1,4,3,2,5,9])
```

Out[4]: 3.375

### 2. Using a Dictionary keep track of the count of numbers (or items) from a list

```
In [44]: def counts(my_list):  
         counts = dict()  
         for item in my_list:  
             counts[item] = my_list.count(item)  
  
         return counts  
  
counts([1,2,1,4,3,2,5,9])
```

```
Out[44]: {1: 2, 2: 2, 4: 1, 3: 1, 5: 1, 9: 1}
```

**3. Using the `counts()` function you created above and the `.split()` function, return a dictionary of most occurring words from the following paragraph. Bonus, remove punctuation from words.**

```
In [48]: paragraph_text = '''
For a minute or two she stood looking at the house, and wondering what
The Fish-Footman began by producing from under his arm a great letter,
Then they both bowed low, and their curls got entangled together.
Alice laughed so much at this, that she had to run back into the wood
Alice went timidly up to the door, and knocked.
'There's no sort of use in knocking,' said the Footman, 'and that for
'Please, then,' said Alice, 'how am I to get in?'
'There might be some sense in your knocking,' the Footman went on with
'I shall sit here,' the Footman remarked, 'till tomorrow-'
At this moment the door of the house opened, and a large plate came sk

import re
new_para = re.sub(r'^a-zA-Z0-9 ', '', paragraph_text)
#print(new_para)
para_list = new_para.split()
#print(para_list)

counts(para_list)
```

```
Out[48]: {'For': 3,
'a': 15,
'minute': 1,
'or': 2,
'two': 2,
'she': 6,
'stood': 1,
'looking': 2,
'at': 6,
'the': 32,
'house': 2,
'and': 17,
'wondering': 1,
'what': 2,
'to': 15,
'do': 1,
'next': 2,
'when': 2,
'suddenly': 1,
'footman': 2}
```

#### 4. Read in a file using `open()` and iterated through the file line-by-line write each line from the file to a new file in a `title()`-ized. Create your own file for input

This is the first line -> This Is The First Line

Hint: There's a function to do this

```
In [34]: file = open("week2-dummy.txt", "r")  
print(file.read())
```

Magna Ultrices Ligula Inceptos Nullam Cras  
Non Porttitor Dis Felis Non Vivamus Est  
Rhoncus per ipsum taciti porta felis, lacus praesent, augue augue. Ne  
c. Lacus. Magna faucibus interdum taciti erat. Euismod iaculis turpis  
eu metus. Tristique convallis porttitor elit eget ante dis curabitur.  
Pellentesque Lacinia Massa Tempor  
Lectus in purus habitant massa taciti aliquam diam est vestibulum nul  
la. Massa. Nec tortor. Montes curabitur. Odio proin placerat magna li  
gula pharetra sagittis habitant vulputate integer rhoncus, est sodale  
s dapibus.  
Cubilia enim sollicitudin curabitur curae;. Varius nibh sem nibh nisi  
cum ultrices ut amet. Elit, molestie ultrices mus cubilia. Condimentu  
m. Consectetuer Etiam, quis feugiat magnis nam eu, gravida pulvinar u  
t accumsan adipiscing quam commodo enim elit tellus morbi condimentum  
erat.

```
In [33]: file = open("week2-dummy.txt", "r")  
newfile = open("week2-dummytitle.txt", "a")  
for x in file:  
    newfile = open("week2-dummytitle.txt", "a")  
    newfile.write(x.title())  
    print(x.title())
```

Magna Ultrices Ligula Inceptos Nullam Cras

Non Porttitor Dis Felis Non Vivamus Est

Rhoncus Per Ipsum Taciti Porta Felis, Lacus Praesent, Augue Augue. Ne  
c. Lacus. Magna Faucibus Interdum Taciti Erat. Euismod Iaculis Turpis  
Eu Metus. Tristique Convallis Porttitor Elit Eget Ante Dis Curabitur.

Pellentesque Lacinia Massa Tempor

Lectus In Purus Habitant Massa Taciti Aliquam Diam Est Vestibulum Nul  
la. Massa. Nec Tortor. Montes Curabitur. Odio Proin Placerat Magna Li  
gula Pharetra Sagittis Habitant Vulputate Integer Rhoncus, Est Sodale  
s Dapibus.

Cubilia Enim Sollicitudin Curabitur Curae;. Varius Nibh Sem Nibh Nisi  
Cum Ultrices Ut Amet. Elit, Molestie Ultrices Mus Cubilia. Condimentu  
m. Consectetuer Etiam, Quis Feugiat Magnis Nam Eu, Gravida Pulvinar U  
t Accumsan Adipiscing Quam Commmodo Enim Elit Tellus Morbi Condimentum  
Erat.

```
In [36]: newfile = open("week2-dummytitle.txt", "r")
print(newfile.read())
```

Magna Ultrices Ligula Inceptos Nullam Cras  
Non Porttitor Dis Felis Non Vivamus Est  
Rhoncus Per Ipsum Taciti Porta Felis, Lacus Praesent, Augue Augue. Ne  
c. Lacus. Magna Faucibus Interdum Taciti Erat. Euismod Iaculis Turpis  
Eu Metus. Tristique Convallis Porttitor Elit Eget Ante Dis Curabitur.  
Pellentesque Lacinia Massa Tempor  
Lectus In Purus Habitant Massa Taciti Aliquam Diam Est Vestibulum Nul  
la. Massa. Nec Tortor. Montes Curabitur. Odio Proin Placerat Magna Li  
gula Pharetra Sagittis Habitant Vulputate Integer Rhoncus, Est Sodale  
s Dapibus.  
Cubilia Enim Sollicitudin Curabitur Curae;. Varius Nibh Sem Nibh Nisi  
Cum Ultrices Ut Amet. Elit, Molestie Ultrices Mus Cubilia. Condimentu  
m. Consectetuer Etiam, Quis Feugiat Magnis Nam Eu, Gravida Pulvinar U  
t Accumsan Adipiscing Quam Commoda Enim Elit Tellus Morbi Condimentum  
Erat.

## Numpy

### 1. Given a list, find the average using a numpy function.

```
In [66]: simple_list = [1,2,1,4,3,2,5,9]
np.mean(simple_list)
```

Out[66]: 3.375

### 2. Given two lists of Heights and Weights of individual, calculate the BMI of those individuals, without writing a for-loop

```
In [103]: heights = [174, 173, 173, 175, 171]
           weights = [88, 83, 92, 74, 77]
           #type(heights)

           heights_format = np.array(heights)/100
           heights_format
           #bmi is weight/height^2

           heightssq = heights_format**2
           heightssq

           bmi = np.array(weights)/heightssq
           bmi
```

```
Out[103]: array([29.06592681, 27.73229978, 30.73941662, 24.16326531, 26.33288875])
```

### 3. Create an array of length 20 filled with random values (between 0 to 1)

```
In [67]: new_array = np.random.randint(2, size=20)
           new_array
```

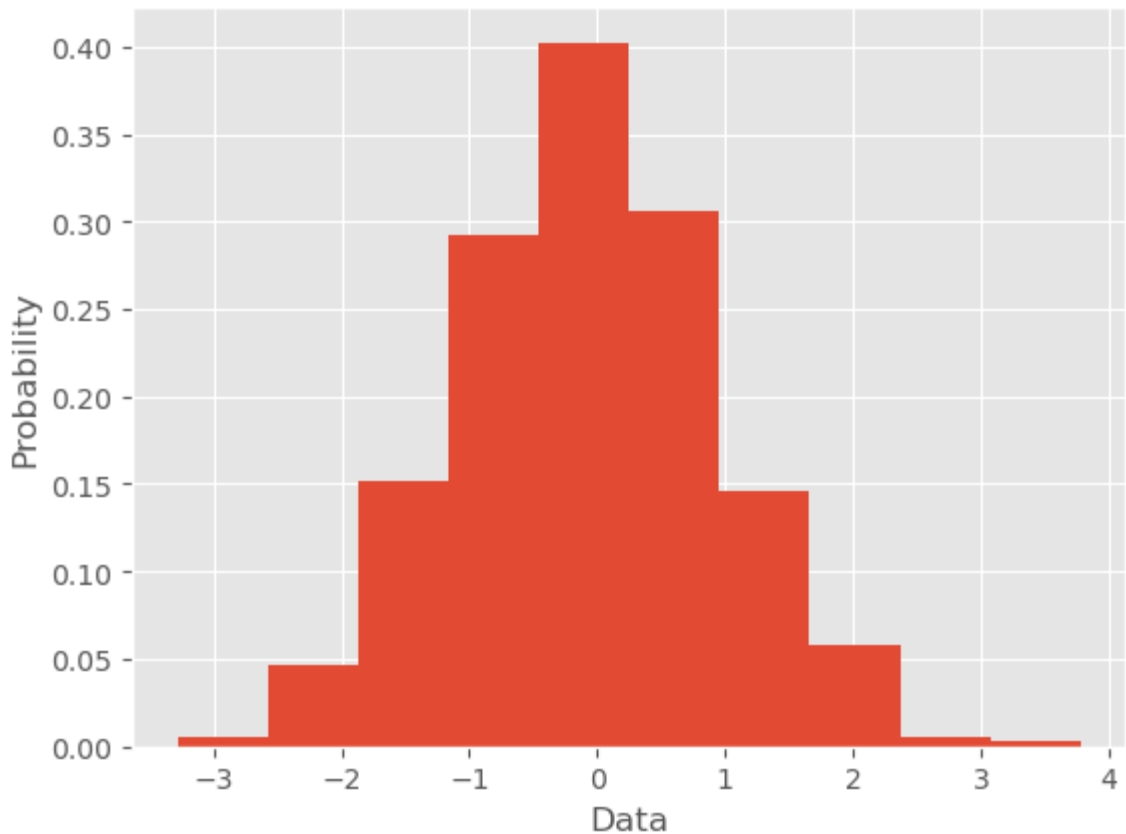
```
Out[67]: array([0, 0, 1, 1, 0, 1, 0, 0, 0, 1, 1, 0, 0, 1, 1, 1, 1, 1, 1])
```

### 4. Create an array with at least 1000 random numbers from normal distributions (normal). Then, plot a histogram of these values (plt.hist).

```
In [16]: random_array = np.random.normal(size=1000)
random_array

plt.hist(random_array, density = True)
plt.ylabel('Probability')
plt.xlabel('Data')
```

```
Out[16]: Text(0.5, 0, 'Data')
```



## Pandas

### 1. Read in a CSV () and display all the columns and their respective data types

```
In [60]: hw_data = pd.read_csv("hw_data.csv", dtype = {'id': int, 'sex': str, '
print(hw_data)
```

	id	sex	weight	height
0	1	M	190	77
1	2	F	120	70
2	3	F	110	68
3	4	M	150	72
4	5	O	120	66
5	6	M	120	60
6	7	F	140	70

## 2. Find the average weight

```
In [61]: hw_data['weight'].mean()
```

```
Out[61]: 135.71428571428572
```

## 3. Find the Value Counts on column sex

```
In [65]: hw_data.groupby(['sex']).count()
```

```
Out[65]:
```

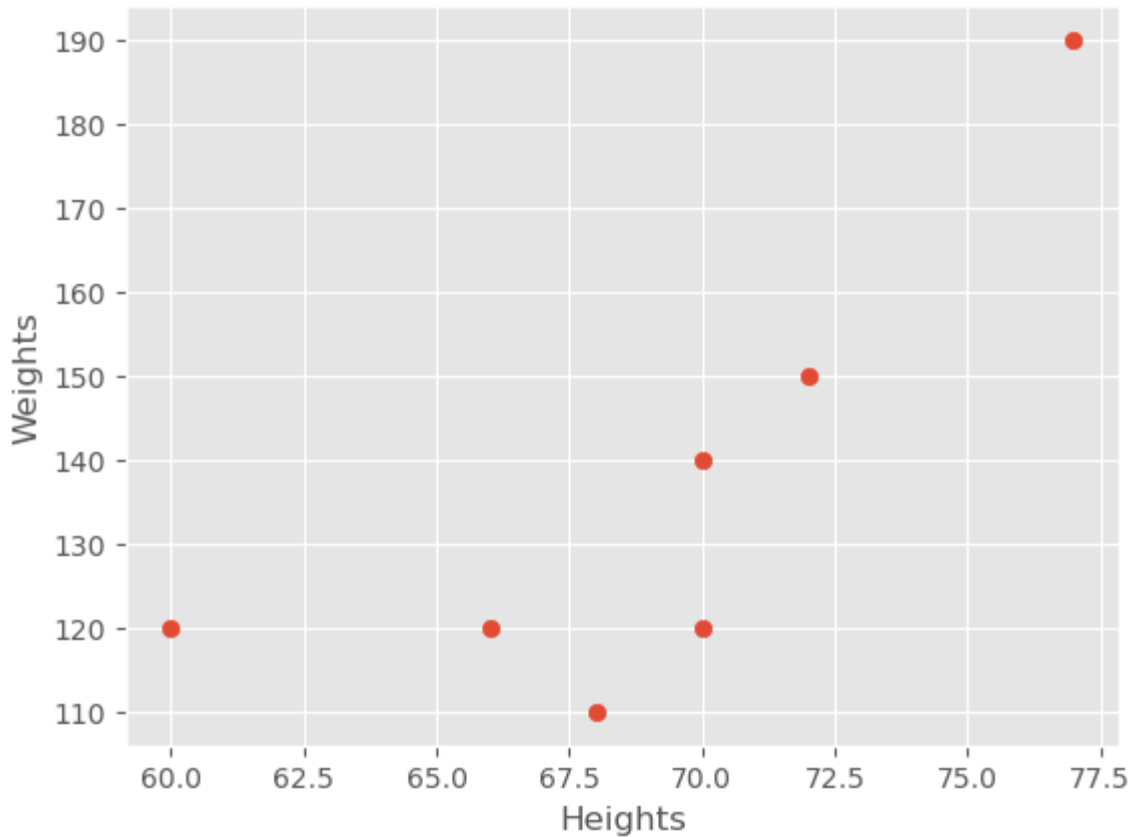
	id	weight	height
sex			
F	3	3	3
M	3	3	3
O	1	1	1

## 4. Plot Height vs. Weight



```
In [107]: plt.scatter(hw_data['height'], hw_data['weight'])  
plt.xlabel('Heights')  
plt.ylabel('Weights')
```

Out[107]: Text(0, 0.5, 'Weights')



## 5. Calculate BMI and save as a new column

```
In [113]: bmi = (hw_data['weight']/(hw_data['height']**2))*703  
hw_data['BMI'] = bmi  
hw_data
```

Out[113]:

	id	sex	weight	height	BMI
0	1	M	190	77	22.528251
1	2	F	120	70	17.216327
2	3	F	110	68	16.723616
3	4	M	150	72	20.341435
4	5	O	120	66	19.366391
5	6	M	120	60	23.433333
6	7	F	140	70	20.085714

## 6. Save sheet as a new CSV file hw\_dataB.csv

```
In [114]: hw_data.to_csv('hw_dataB.csv')
```

### Run the following (Mac)

```
In [115]: !cat hw_dataB.csv
```

```
,id,sex,weight,height,BMI
0,1,M,190,77,22.52825096980941
1,2,F,120,70,17.216326530612243
2,3,F,110,68,16.72361591695502
3,4,M,150,72,20.341435185185187
4,5,0,120,66,19.366391184573004
5,6,M,120,60,23.433333333333334
6,7,F,140,70,20.085714285714285
```

### Run the following (Windows)

```
In [ ]: !type hw_dataB.csv
```