Lab 11

Jennifer Lin jenniferyjlin@berkeley.edu

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import math
```

1

Read in the file "https://raw.githubusercontent.com/uiuc-cse/data-fa14/gh-pages/data/iris.csv" into a data frame. This is the same dataset we worked with on R.

```
iris = pd.read_csv('https://raw.githubusercontent.com/uiuc-cse/data-fa14/gh-pages/data/iris.csv')
iris.head() # check the input format
```

	sepal_length	sepal_width	petal_length	petal_width	species
0	5.1	3.5	1.4	0.2	setosa
1	4.9	3.0	1.4	0.2	setosa
2	4.7	3.2	1.3	0.2	setosa
3	4.6	3.1	1.5	0.2	setosa
4	5.0	3.6	1.4	0.2	setosa

2

Create a new data frame only with the columns Sepal.Length, Sepal.Width. Write it into an Excel file.

```
iris_sepal = iris[['sepal_length','sepal_width']]
iris_sepal.head()
```

	sepal_length	sepal_width
0	5.1	3.5
1	4.9	3.0
2	4.7	3.2
3	4.6	3.1
4	5.0	3.6

```
iris_sepal.to_excel("output.xlsx")
```

3

Create 3 new data frames for each of the species.

```
# group
iris_grouped = iris.groupby(iris.species)
# 3 data frmaes
iris_setosa = iris_grouped.get_group("setosa")
iris_versicolor = iris_grouped.get_group("versicolor")
iris_virginica = iris_grouped.get_group("virginica")

# view the head of the iris_setosa
iris_setosa.head()
```

	sepal_length	sepal_width	petal_length	petal_width	species
0	5.1	3.5	1.4	0.2	setosa
1	4.9	3.0	1.4	0.2	setosa
2	4.7	3.2	1.3	0.2	setosa
3	4.6	3.1	1.5	0.2	setosa
4	5.0	3.6	1.4	0.2	setosa

```
# view the head of the iris_versicolor
iris_versicolor.head()
```

	sepal_length	sepal_width	petal_length	petal_width	species
50	7.0	3.2	4.7	1.4	versicolor
51	6.4	3.2	4.5	1.5	versicolor
52	6.9	3.1	4.9	1.5	versicolor
53	5.5	2.3	4.0	1.3	versicolor
54	6.5	2.8	4.6	1.5	versicolor

```
# view the head of the iris_virginica
iris_virginica.head()
```

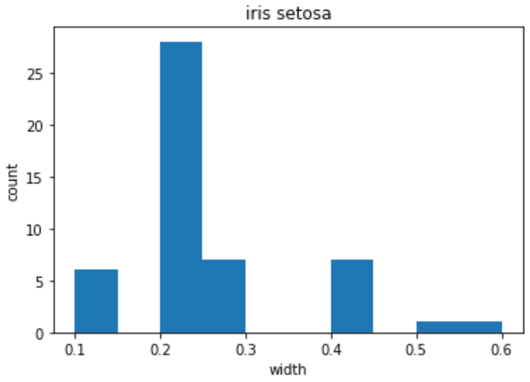
	sepal_length	sepal_width	petal_length	petal_width	species
100	6.3	3.3	6.0	2.5	virginica
101	5.8	2.7	5.1	1.9	virginica
102	7.1	3.0	5.9	2.1	virginica
103	6.3	2.9	5.6	1.8	virginica
104	6.5	3.0	5.8	2.2	virginica

4

Create a histogram of Pedal.Width for each of the 3 species. hint: use numpy.hist()

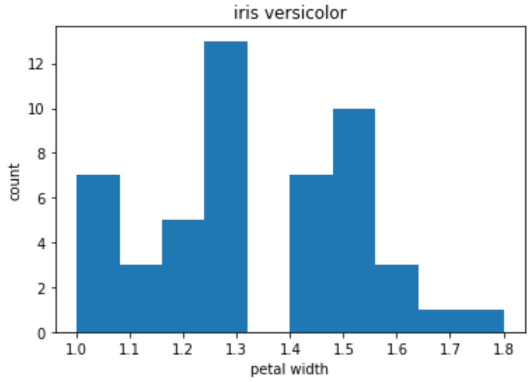
```
plt.hist(iris_setosa['petal_width'])
plt.title("iris setosa")
plt.xlabel("width")
plt.ylabel("count")
# plt.hist(iris_setosa['petal_width'], bins=np.arange(0, 1, 0.05).tolist())
```

```
Text(0, 0.5, 'count')
```



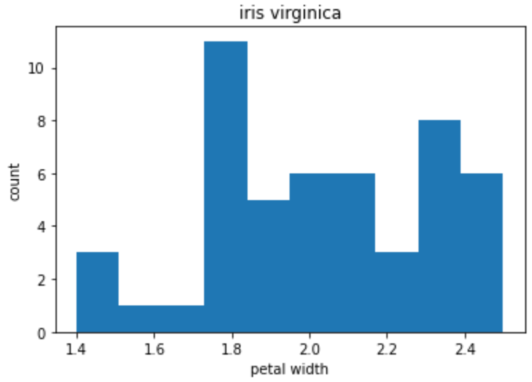
```
plt.hist(iris_versicolor['petal_width'])
plt.title("iris versicolor")
plt.xlabel("petal width")
plt.ylabel("count")
```

```
Text(0, 0.5, 'count')
```



```
plt.hist(iris_virginica['petal_width'])
plt.title("iris virginica")
plt.xlabel("petal width")
plt.ylabel("count")
```

```
Text(0, 0.5, 'count')
```



5

Create a new data frame where you perform some sort of transform on a column of numerical values in maximum three lines of code (e.g. multiplying by 2, taking the logarithm).

```
iris_transform = iris['sepal_length']*3*np.log10(iris['sepal_length'])/2*iris['sepal_length'].sum()/iris['sepal_length'
iris_transform.head()
```

```
0    600.559144
1    562.839614
2    525.710265
3    507.374678
4    581.627097
Name: sepal_length, dtype: float64
```

6

Create a list data structure from the column Sepal.Length and write a function with the input as that list which returns the mean of the column

```
# create a list data structure from the column Sepal.Length
iris_sepal_length_list = iris['sepal_length']
# write a function with the input as that list which returns the mean of the column
def mean_fun(input_list):
  output_mean = np.mean(input_list)
  return round(output_mean,4)
# mean of the column Sepal.Length
print(mean_fun(iris_sepal_length_list))
```

```
5.8433
```

7

Create a dictionary with the keys being each column in the data frame (except for species) and the value as the mean of each column using th function you wrote above. hint: you need to use the mean function in numpy since it is not a built-in keyword in Python

```
column_mean = {}
for i in range(4):
  colname_i = iris.columns[i]
  column_mean[colname_i] = mean_fun(iris[colname_i])
print(column_mean)
```

```
{'sepal_length': 5.8433, 'sepal_width': 3.054, 'petal_length': 3.7587, 'petal_width': 1.1987}
```