

151_15_5116_LAB2

February 6, 2018

```
In [1]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import matplotlib.patches as mp
```

```
In [2]: df = pd.read_csv('iris.csv')
df.head(5)
```

```
Out[2]:
```

	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

```
In [3]: df.tail(3)
```

```
Out[3]:
```

	sepal_length	sepal_width	petal_length	petal_width	species
147	6.5	3.0	5.2	2.0	virginica
148	6.2	3.4	5.4	2.3	virginica
149	5.9	3.0	5.1	1.8	virginica

```
In [4]: df.sort_values('petal_length').head()
```

```
Out[4]:
```

	sepal_length	sepal_width	petal_length	petal_width	species
22	4.6	3.6	1.0	0.2	setosa
13	4.3	3.0	1.1	0.1	setosa
14	5.8	4.0	1.2	0.2	setosa
35	5.0	3.2	1.2	0.2	setosa
36	5.5	3.5	1.3	0.2	setosa

```
In [9]: new_df = df.groupby(['species'])[0:4]['sepal_length', 'sepal_width', 'petal_length', 'petal_width']
print(type(new_df))
```

```
<class 'pandas.core.frame.DataFrame'>
```

```
In [10]: print(new_df)
```

	sepal_length	sepal_width	petal_length	petal_width
species				
setosa	5.006	3.418	1.464	0.244
versicolor	5.936	2.770	4.260	1.326
virginica	6.588	2.974	5.552	2.026

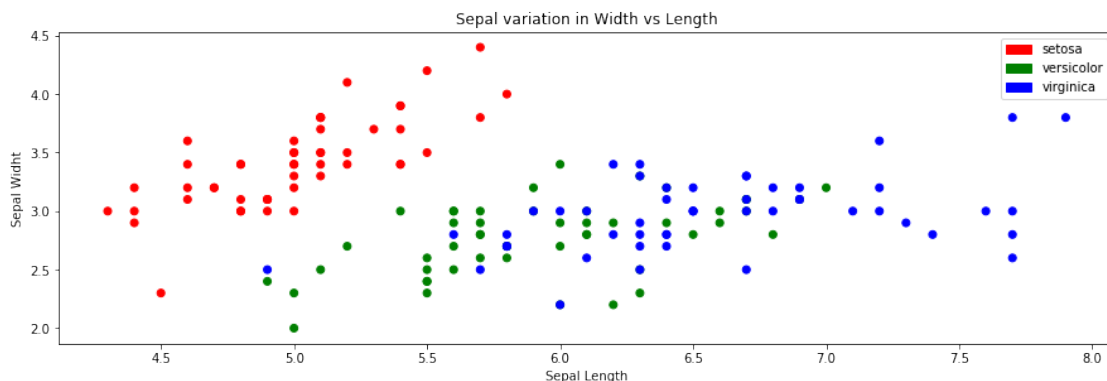
```
In [11]: sepal_length=df['sepal_length']
sepal_width= df['sepal_width']
petal_length = df['sepal_width']
petal_width = df['petal_width']
species = df['species']
```

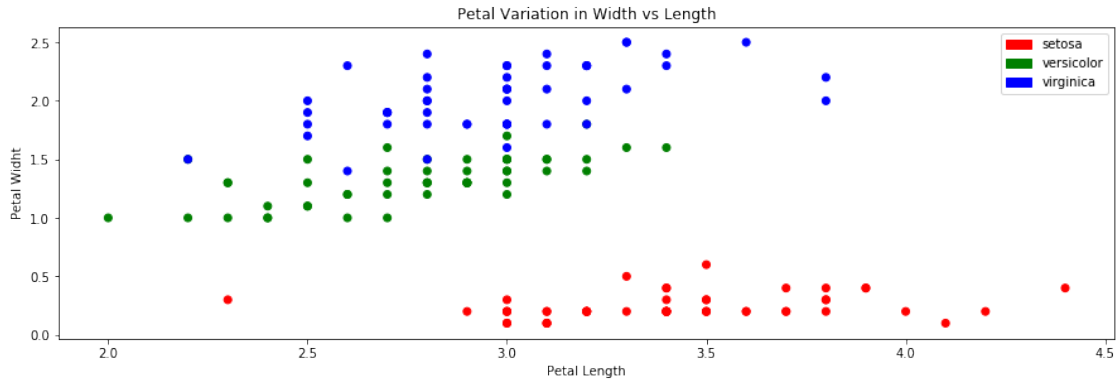
```
In [12]: pairs={'setosa' : 'r', 'versicolor' : 'g', 'virginica' : 'b'}
labels = [mp.Patch(color=cl, label=la) for la, cl in pairs.items()]
```

```
In [13]: plt.figure(figsize=(15,10))
plt.subplot(2, 1,2)
plt.scatter(sepal_length, sepal_width, c=[pairs[i] for i in species], label=[pairs[i] f
plt.ylabel('Sepal Widht')
plt.xlabel('Sepal Length')
plt.title('Sepal variation in Width vs Length')
plt.legend(handles = labels)
```

```
Out[13]: <matplotlib.legend.Legend at 0x7f9ce6473e80>
```

```
In [14]: plt.figure(figsize=(15,10))
plt.subplot(2, 1, 2)
plt.scatter(petal_length, petal_width, c=[pairs[i] for i in species], label=[pairs[i] f
plt.ylabel('Petal Widht')
plt.xlabel('Petal Length')
plt.title('Petal Variation in Width vs Length')
plt.legend(handles = labels)
plt.show()
```





```
In [15]: def check(x):
          y = []
          for i in range(len(x)):
              if df['sepal_length'][i] < 5:
                  y.append(0)
              else:
                  y.append(1)
          return y

In [16]: df['Calyx Width'] = check(df['sepal_length'])
          df['Calyx Width'] = df.apply(lambda df: 0 if df['sepal_length'] < 5 else 1, axis=1)

In [17]: df.head()

Out[17]:
```

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

```
In [18]: plt.figure(figsize=(10,5))
          plt.hist(sepal_length, bins=25,color='r')
          plt.ylabel('Number')
          plt.xlabel('Sepal Length')
          plt.title('Sepal Length Histogram')
          plt.show()
```

