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

In [3]:
```python
df.head(5)
```

Out[3]:

	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 [4]:
```python
df.groupby('species')['sepal_length','sepal_width','petal_length','petal_width'].
df.tail(3)
```

Out[4]:

	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 [ ]:
```python
#task 2
```

In [5]:
```python
df.sort_values('petal_length').head()
```

Out[5]:

	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 [6]:
```python
new_df = df.groupby(['species'][0:4])['sepal_length','sepal_width','petal_length'
print(type(new_df))
```

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

```
In [7]: #task 3
        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
```

```
#task 4
```

```
In [8]:
        sepal_length=df['sepal_length']
        sepal_width= df['sepal_width']
        petal_length = df['sepal_width']
        petal_width = df['petal_width']
        species = df['species']
```

In [9]:
```python
pairs={'setosa' :'r','versicolor' :'g','virginica' :'b'}
labels = [mp.Patch(color=cl, label=la) for la, cl in pairs.items()]

plt.figure(figsize=(15,10))
plt.subplot(2, 1,2)
plt.scatter(sepal_length, sepal_width, c=[pairs[i] for i in species], label=[pair
plt.ylabel('Sepal Widht')                    # set y label
plt.xlabel('Sepal Length')                   # set x label
plt.title('Sepal variation in Width vs Length')  # give it a title
plt.legend(handles = labels)


plt.figure(figsize=(15,10))
plt.subplot(2, 1, 2)
plt.scatter(petal_length, petal_width, c=[pairs[i] for i in species], label=[pair
plt.ylabel('Petal Widht')                    # set y label
plt.xlabel('Petal Length')                   # set x label
plt.title('Petal Variation in Width vs Length')  # give it a title
plt.legend(handles = labels)
plt.show()
```
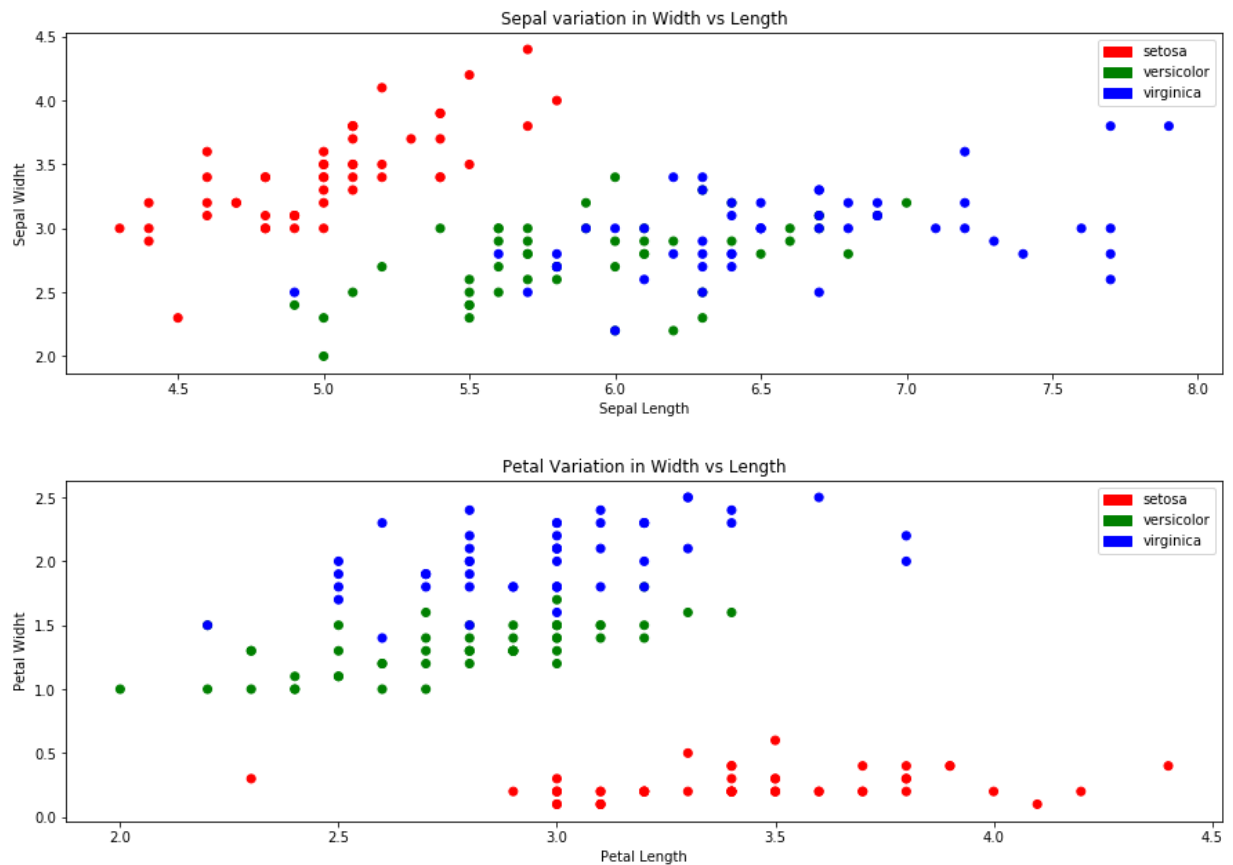


```
#task 5
```

```
In [10]: 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 [11]: df['Calyx Width'] =  check(df['sepal_length'])

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

```
In [12]: df.head()
```

Out[12]:

	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 [13]: plt.figure(figsize=(10,5))              # set fig size
         plt.hist(sepal_length, bins=25,color='r')   # Set histogram
         plt.ylabel('Number')                    # set y label
         plt.xlabel('Sepal Length')              # set x label
         plt.title('Sepal Length Histogram')     # give it a title
         plt.show()
```