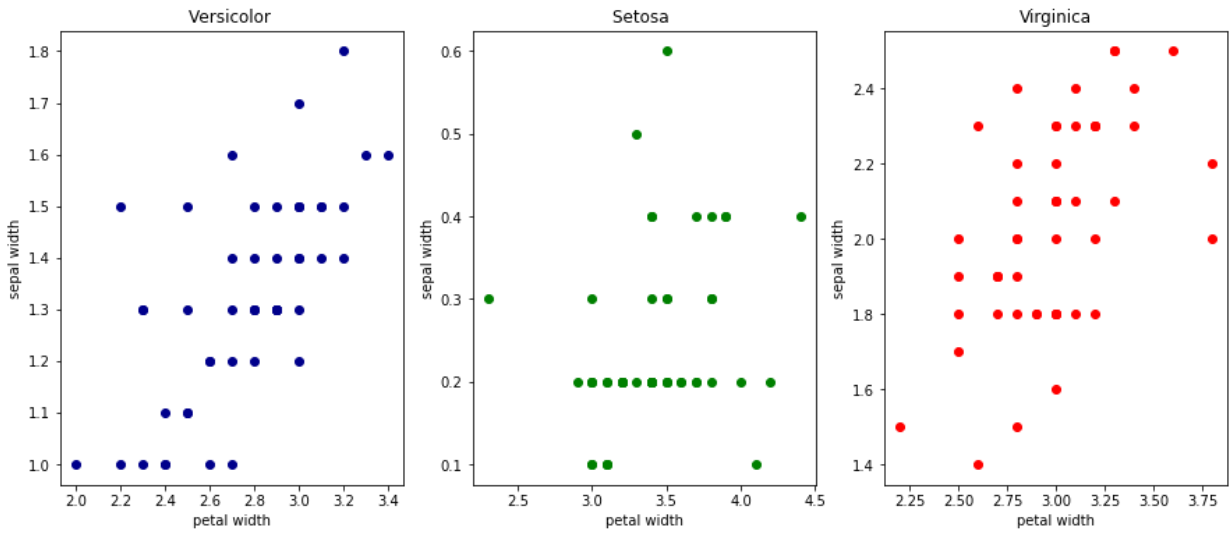
In [1]:
```
#Importing libraries we need
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
import math
```

In [2]:
```
#import and read data
irisData = pd.read_csv('iris.csv')
```

In [42]:
```
#petal and sepal width by species

veSepWidthSet = irisData[irisData['species']=='versicolor'].petal_width
vePetWidthSet = irisData[irisData['species']=='versicolor'].sepal_width

seSepWidthSet = irisData[irisData['species']=='setosa'].petal_width
sePetWidthSet = irisData[irisData['species']=='setosa'].sepal_width

viSepWidthSet = irisData[irisData['species']=='virginica'].petal_width
viPetWidthSet = irisData[irisData['species']=='virginica'].sepal_width

#petal and sepal length by species

veSepLengthSet = irisData[irisData['species']=='versicolor'].petal_length
vePetLengthSet = irisData[irisData['species']=='versicolor'].sepal_length

seSepLengthSet = irisData[irisData['species']=='setosa'].petal_length
sePetLengthSet = irisData[irisData['species']=='setosa'].sepal_length

viSepLengthSet = irisData[irisData['species']=='virginica'].petal_length
viPetLengthSet = irisData[irisData['species']=='virginica'].sepal_length
```

In [34]:
```
#We can see that there is a positive correlation between 2 versicolor petal


fig, (ax1,ax2,ax3) = plt.subplots(nrows = 1,ncols= 3,figsize=(15, 6))

ax1.scatter(vePetWidthSet, veSepWidthSet,c='darkblue')
ax1.set_title('Versicolor ')
ax1.set_xlabel('petal width')
ax1.set_ylabel('sepal width')

ax2.scatter(sePetWidthSet, seSepWidthSet,c='green')
ax2.set_title('Setosa ')
ax2.set_xlabel('petal width')
ax2.set_ylabel('sepal width')

ax3.scatter(viPetWidthSet, viSepWidthSet,c='red')
ax3.set_title('Virginica')
ax3.set_xlabel('petal width')
ax3.set_ylabel('sepal width')
```

Out[34]:
```
Text(0, 0.5, 'sepal width')
```



In [39]:
```
print("The correlation coefficient on Virginica is equal:\n", round(viSepWi
```

```
The correlation coefficient on Versicolor is equal:
 0.54
```

In [40]:
```
print("The correlation coefficient on Versicolor is equal:\n",round(vePetWi
```

```
The correlation coefficient on Versicolor is equal:
 0.66
```

In [41]:
```
print("The correlation coefficient on Virginica is equal:\n", round(sePetWi
```

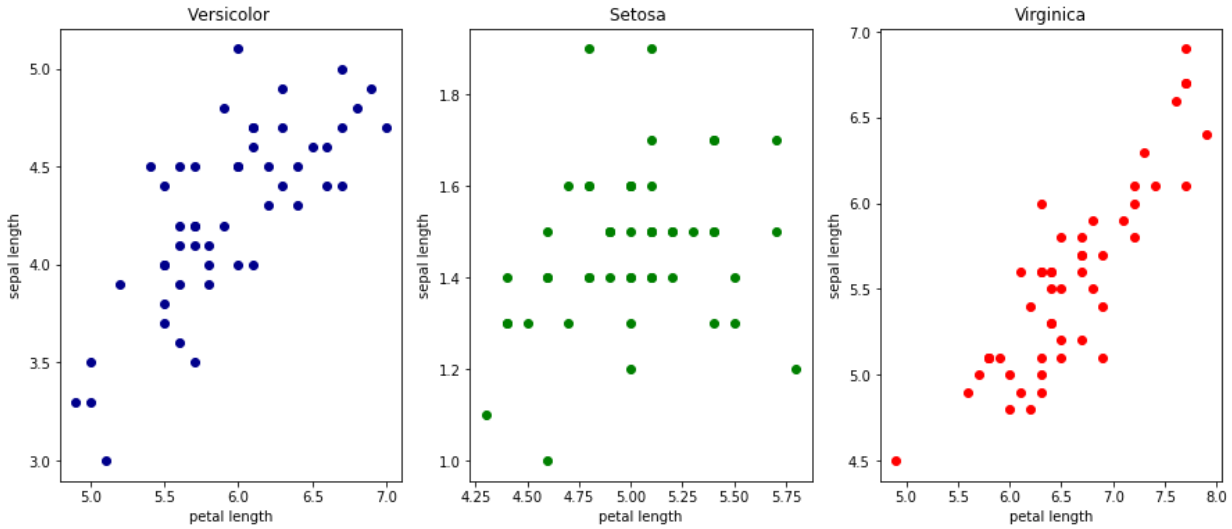
```
The correlation coefficient on Virginica is equal:
 0.28
```

Based on thenumbers above, I would conclude that petal and sepal

width are more correlated on Versicolor than Virginica species while it is almost non existent on Setosa given the correlation error from the best fit line

In [43]:
```
fig, (ax1,ax2,ax3) = plt.subplots(nrows = 1,ncols= 3,figsize=(15, 6))

ax1.scatter(vePetLengthSet, veSepLengthSet,c='darkblue')
ax1.set_title('Versicolor ')
ax1.set_xlabel('petal length')
ax1.set_ylabel('sepal length')

ax2.scatter(sePetLengthSet, seSepLengthSet,c='green')
ax2.set_title('Setosa ')
ax2.set_xlabel('petal length')
ax2.set_ylabel('sepal length')

ax3.scatter(viPetLengthSet, viSepLengthSet,c='red')
ax3.set_title('Virginica')
ax3.set_xlabel('petal length')
ax3.set_ylabel('sepal length')
```

Out[43]: Text(0, 0.5, 'sepal length')



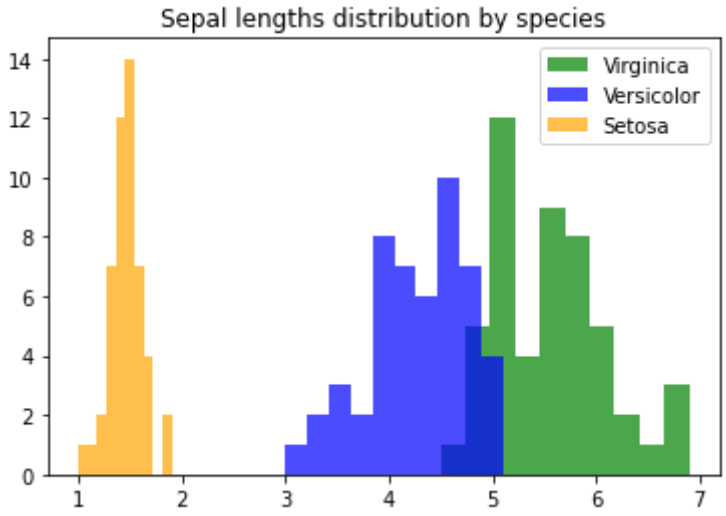
In [46]:
```
print("Virginica:\n",round(viSepLengthSet.corr(viPetLengthSet),2))
print("Setosa:\n",round(seSepLengthSet.corr(sePetLengthSet),2))
print("Versicolor:\n",round(veSepLengthSet.corr(vePetLengthSet),2))
```

```
Virginica:
 0.86
Setosa:
 0.26
Versicolor:
 0.75
```

Based on thenumbers above, it is fair to say that petal and sepal length are strongly correlated on Virginica species,strong on Versicolor while weak or non existent on Setosa

In [127]:

```
# plotting second histogram
plt.hist(viSepLengthSet, label='Virginica', alpha=0.7,color='green')
plt.hist(veSepLengthSet, label='Versicolor', alpha=0.7, color='blue')
plt.hist(seSepLengthSet, label='Setosa', alpha=0.7,color='orange')
plt.legend()
plt.title("Sepal lengths distribution by species")
plt.show()
```



End