In [2]:

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt

#we want to se the data in jupyter

df = pd.read_csv("iris_csv.csv")
df.head()
```

Out[2]:

	sepallength	sepalwidth	petallength	petalwidth
0	5.1	3.5	1.4	0.2
1	4.9	3.0	1.4	0.2
2	4.7	3.2	1.3	0.2
3	4.6	3.1	1.5	0.2
4	5.0	3.6	1.4	0.2

In [3]:

```
#overall data performance with entries
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 150 entries, 0 to 149
Data columns (total 5 columns):
sepallength    150 non-null float64
sepalwidth     150 non-null float64
petallength    150 non-null float64
petalwidth     150 non-null float64
class          150 non-null object
dtypes: float64(4), object(1)
memory usage: 5.9+ KB
```

In [4]:

```
#we want to see the array dimension
df.shape
```

Out[4]:

```
(150, 5)
```

In [5]:

```
#statistics about the dataset
df.describe()
```

Out[5]:

	sepallength	sepalwidth	petallength	petal
count	150.000000	150.000000	150.000000	150.0(
mean	5.843333	3.054000	3.758667	1.1(
std	0.828066	0.433594	1.764420	0.7(
min	4.300000	2.000000	1.000000	0.1(
25%	5.100000	2.800000	1.600000	0.3(
50%	5.800000	3.000000	4.350000	1.3(
75%	6.400000	3.300000	5.100000	1.8(
max	7.900000	4.400000	6.900000	2.5(

In [6]:

```
#checking for  null values

df.isnull().sum()
```

Out[6]:

```
sepallength    0
sepalwidth     0
petallength    0
petalwidth     0
class          0
dtype: int64
```

In [24]:

```
#we sre going to do univariate analysis.
df.groupby('class').agg(['mean','median','quantil
e'])
```

Out[24]:

	sepallength			sepalwidth	
	mean	median	quantile	mean	mediar
class					
Iris-setosa	5.006	5.0	5.0	3.418	3.4
Iris-versicolor	5.936	5.9	5.9	2.770	2.8
Iris-virginica	6.588	6.5	6.5	2.974	3.0

In [22]:

```
df.groupby('class').agg([np.mean,np.median])
```

Out[22]:

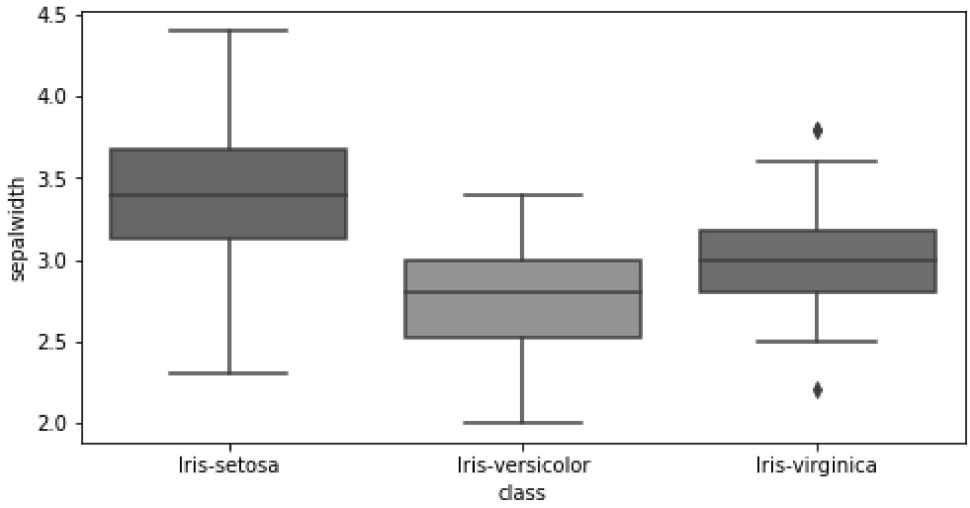
	sepallength		sepalwidth		petaller
	mean	median	mean	median	mean
class					
Iris-setosa	5.006	5.0	3.418	3.4	1.464
Iris-versicolor	5.936	5.9	2.770	2.8	4.260
Iris-virginica	6.588	6.5	2.974	3.0	5.552

In [14]:

```
##box plot
plt.figure(figsize=(8,4))
sns.boxplot(x='class',y ='sepalwidth', data=df, pale
tte=None)
```

Out[14]:

```
<matplotlib.axes._subplots.AxesSubplot
at 0x370c6b0860>
```

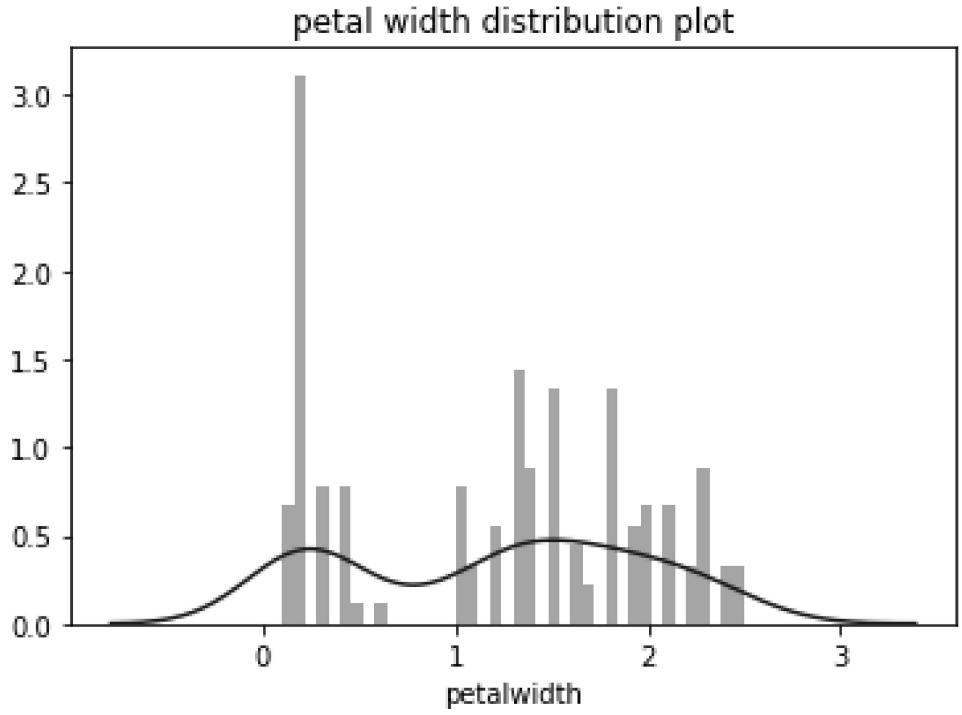


In [26]:

```
#distribution of particular species
sns.distplot(a=df['petalwidth'],bins=40, color='b')
plt.title('petal width distribution plot')
```

Out[26]:

```
Text(0.5, 1.0, 'petal width distributio
n plot')
```

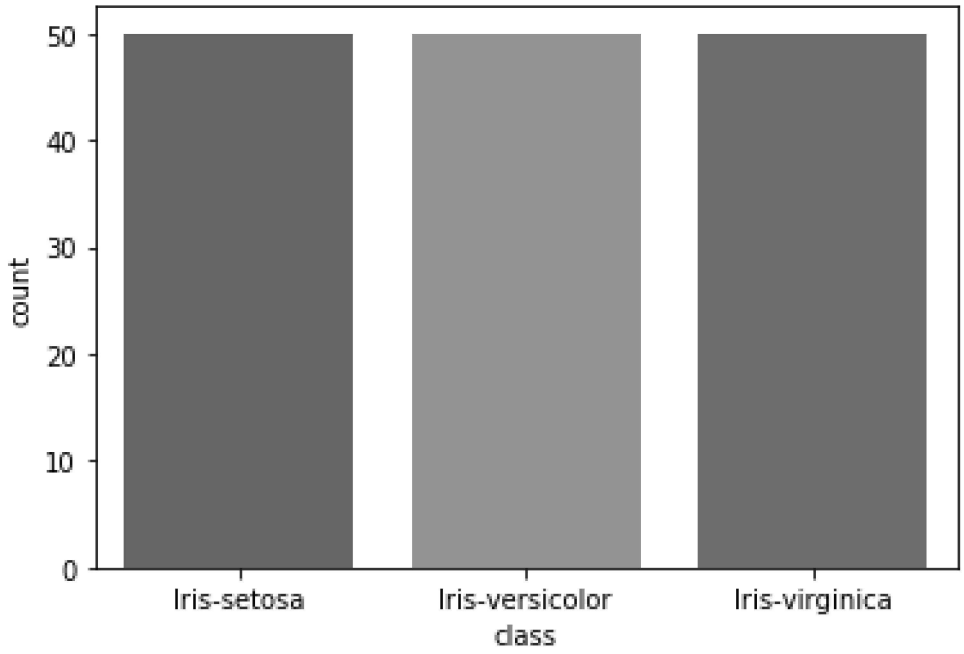


In [27]:

```
#count the number of observation of each species
sns.countplot(x='class', data=df)
```

Out[27]:

```
<matplotlib.axes._subplots.AxesSubplot
at 0x370c9e2cc0>
```

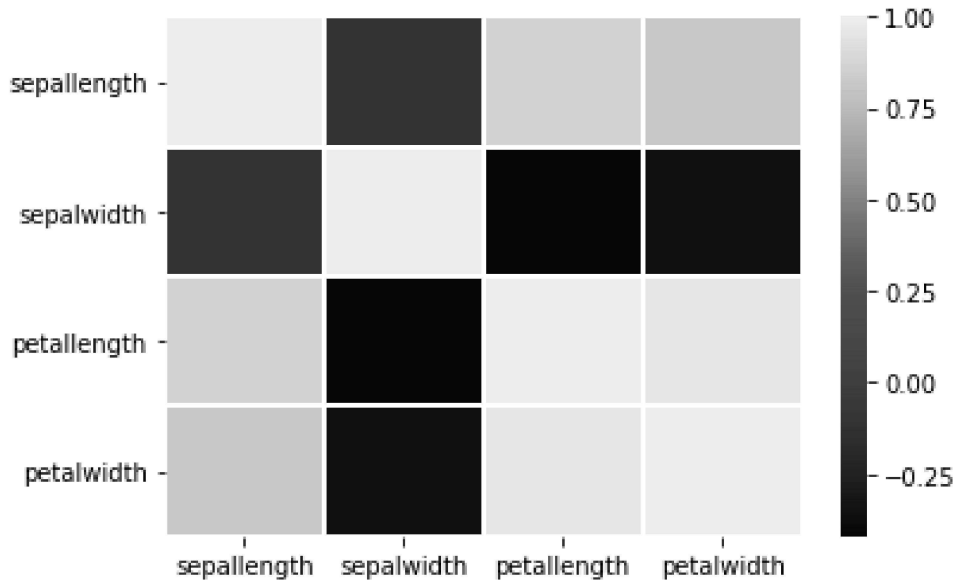


In [28]:

```
##correlation map using heatmap
sns.heatmap(df.corr(), linecolor='white', linewidths
=1)
```

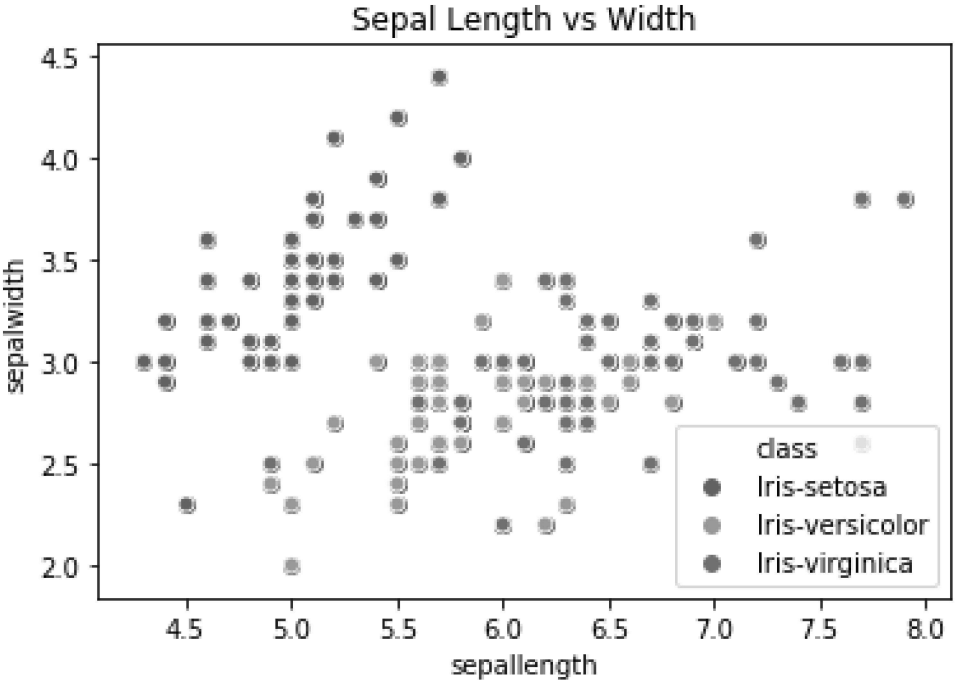
Out[28]:

```
<matplotlib.axes._subplots.AxesSubplot
at 0x370ca63128>
```



In [37]:

```
#multivariate analysis
#analysis between two or more variable of feature
#scatter plot is suitable for the relation between t
wo or more feature
axis = plt.axes()
axis.scatter(df.sepallength, df.sepalwidth)
axis.set(xlabel="Sepal Length (cm)", ylabel="Sepal W
idth (cm)", title ="Sepal Length vs Width");
sns.scatterplot(x="sepallength", y = "sepalwidth", h
ue = "class", data=df)
plt.show()
```



In [ ]: