Using the iris dataset (file on Learning Activities Page) create a summary data plot for the following:

1.Sepal length and width by species 2.Petal length and width by species

Iris Flower Data Set https://en.wikipedia.org/wiki/Iris_flower_data_set (https://en.wikipedia.org/wiki/Iris_flower_data_set)

Richa Patel

Iris data Set 2

Any correlation between sepal and petal width by species? Any correlation between sepal and petal length by species? How are the sepal lengths distributed by species?

In [62]:
```
# import pandas as pd
import numpy as np
from numpy import cov
import matplotlib.pyplot as plt
import seaborn as sns

from scipy.stats import pearsonr

# read file for IRIS Data

iris = pd.read_csv('iris.csv')
print(iris)

print(iris.describe())
iris.info()

iris['species'].value_counts()
```

```
#Correlation with any variables
print(iris.corr())

# 1. correlation between sepal and petal width by species

iris_setosa = iris.loc[iris['species'] == 'setosa']
iris_versicolor = iris.loc[iris['species'] == 'versicolor']
iris_virginica = iris.loc[iris['species'] == 'virginica']

print(iris_setosa)

# scatter plot of sepal and petal width by species
sns.FacetGrid(iris, hue ="species", height =5).map(plt.scatter, 'sepal_width',
 'petal_width').add_legend()


#2. Correlation between sepal and petal length by species
# scatter plot of sepal and petal length by species

sns.FacetGrid(iris, hue ="species", height = 5).map(plt.scatter, 'sepal_length',
 'petal_length').add_legend()


#3. How are the sepal lengths distributed by species?
# To generate a Scatterplot for Sepal_length and Sepal_width using pandas
iris.plot(kind="scatter",x="sepal_length" , y="sepal_width")
plt.show()

# Make change colors and size for display plot
iris.plot(kind="scatter",x="sepal_length",y="sepal_width",color="Red",s=90)
plt.show()

#Scatterplot for Petal_length and Petal_width using pandas
iris.plot(kind="scatter" , x="petal_length",y="petal_width")
plt.show()
sns.boxplot(x="species", y="sepal_length", data=iris)
plt.show()
```

```
sns.boxplot(x="species", y="sepal_width", data=iris)
plt.show()

#we can see there is a clear differences in
#the size of the sepal length with the different species.
#The box plots describe that The virginica species has the longest features in
#sepal_lengths as compared to others.
#The setosa species has the smallest features.


# Petal length and width by species
sns.boxplot(x="species" , y="petal_length", data=iris)
plt.show()

sns.boxplot(x="species" , y="petal_width" , data=iris)
plt.show()
```

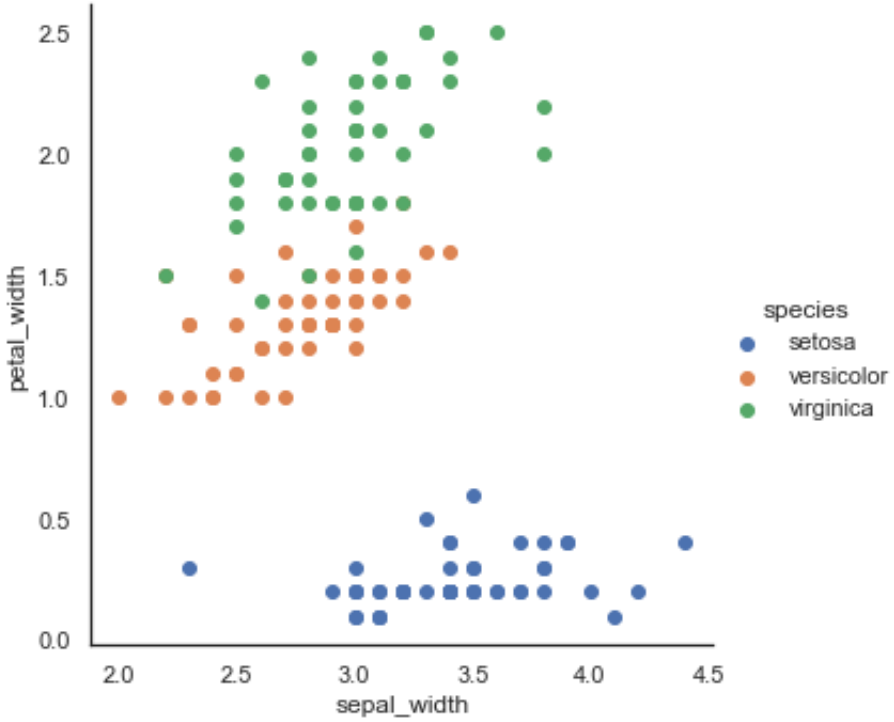
```
     sepal_length  sepal_width  petal_length  petal_width    species
0             5.1          3.5           1.4          0.2     setosa
1             4.9          3.0           1.4          0.2     setosa
2             4.7          3.2           1.3          0.2     setosa
3             4.6          3.1           1.5          0.2     setosa
4             5.0          3.6           1.4          0.2     setosa
..            ...          ...           ...          ...        ...
145           6.7          3.0           5.2          2.3  virginica
146           6.3          2.5           5.0          1.9  virginica
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

[150 rows x 5 columns]
       sepal_length  sepal_width  petal_length  petal_width
count    150.000000   150.000000    150.000000   150.000000
mean       5.843333     3.054000      3.758667     1.198667
std        0.828066     0.433594      1.764420     0.763161
```

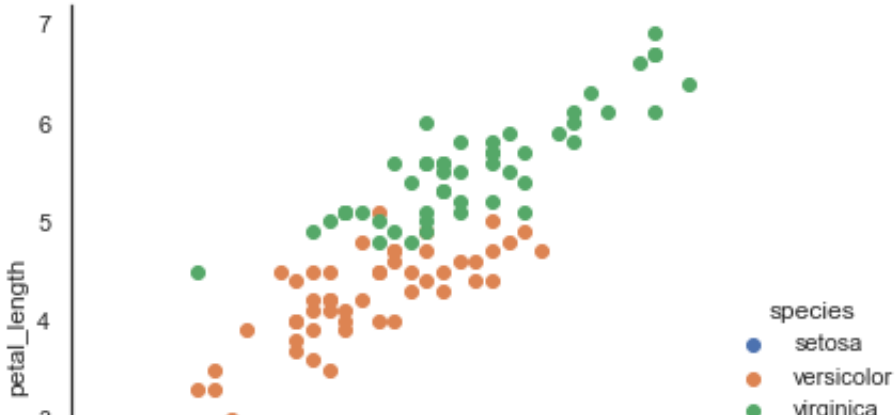
```
min         4.300000     2.000000      1.000000     0.100000
25%         5.100000     2.800000      1.600000     0.300000
50%         5.800000     3.000000      4.350000     1.300000
75%         6.400000     3.300000      5.100000     1.800000
max         7.900000     4.400000      6.900000     2.500000
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 150 entries, 0 to 149
Data columns (total 5 columns):
 #   Column        Non-Null Count  Dtype  
---  ------        --------------  -----  
 0   sepal_length  150 non-null    float64
 1   sepal_width   150 non-null    float64
 2   petal_length  150 non-null    float64
 3   petal_width   150 non-null    float64
 4   species       150 non-null    object 
dtypes: float64(4), object(1)
memory usage: 6.0+ KB
              sepal_length  sepal_width  petal_length  petal_width
sepal_length      1.000000    -0.109369      0.871754     0.817954
sepal_width      -0.109369     1.000000     -0.420516    -0.356544
petal_length      0.871754    -0.420516      1.000000     0.962757
petal_width       0.817954    -0.356544      0.962757     1.000000
    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
5            5.4          3.9           1.7          0.4  setosa
6            4.6          3.4           1.4          0.3  setosa
7            5.0          3.4           1.5          0.2  setosa
8            4.4          2.9           1.4          0.2  setosa
9            4.9          3.1           1.5          0.1  setosa
10           5.4          3.7           1.5          0.2  setosa
11           4.8          3.4           1.6          0.2  setosa
12           4.8          3.0           1.4          0.1  setosa
13           4.3          3.0           1.1          0.1  setosa
```

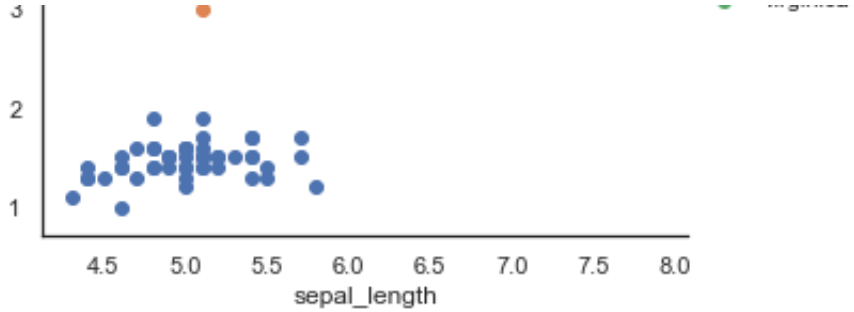
14	5.8	4.0	1.2	0.2	setosa
15	5.7	4.4	1.5	0.4	setosa
16	5.4	3.9	1.3	0.4	setosa
17	5.1	3.5	1.4	0.3	setosa
18	5.7	3.8	1.7	0.3	setosa
19	5.1	3.8	1.5	0.3	setosa
20	5.4	3.4	1.7	0.2	setosa
21	5.1	3.7	1.5	0.4	setosa
22	4.6	3.6	1.0	0.2	setosa
23	5.1	3.3	1.7	0.5	setosa
24	4.8	3.4	1.9	0.2	setosa
25	5.0	3.0	1.6	0.2	setosa
26	5.0	3.4	1.6	0.4	setosa
27	5.2	3.5	1.5	0.2	setosa
28	5.2	3.4	1.4	0.2	setosa
29	4.7	3.2	1.6	0.2	setosa
30	4.8	3.1	1.6	0.2	setosa
31	5.4	3.4	1.5	0.4	setosa
32	5.2	4.1	1.5	0.1	setosa
33	5.5	4.2	1.4	0.2	setosa
34	4.9	3.1	1.5	0.1	setosa
35	5.0	3.2	1.2	0.2	setosa
36	5.5	3.5	1.3	0.2	setosa
37	4.9	3.1	1.5	0.1	setosa
38	4.4	3.0	1.3	0.2	setosa
39	5.1	3.4	1.5	0.2	setosa
40	5.0	3.5	1.3	0.3	setosa
41	4.5	2.3	1.3	0.3	setosa
42	4.4	3.2	1.3	0.2	setosa
43	5.0	3.5	1.6	0.6	setosa
44	5.1	3.8	1.9	0.4	setosa
45	4.8	3.0	1.4	0.3	setosa
46	5.1	3.8	1.6	0.2	setosa
47	4.6	3.2	1.4	0.2	setosa
48	5.3	3.7	1.5	0.2	setosa
49	5.0	3.3	1.4	0.2	setosa

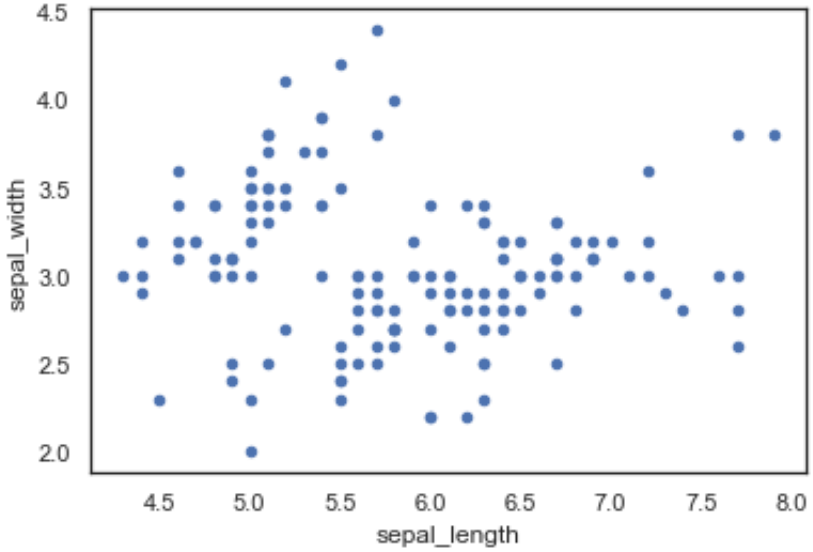
```
...t argument looks like a single numeric RGB or RGBA sequence, which should be avoided as value
```

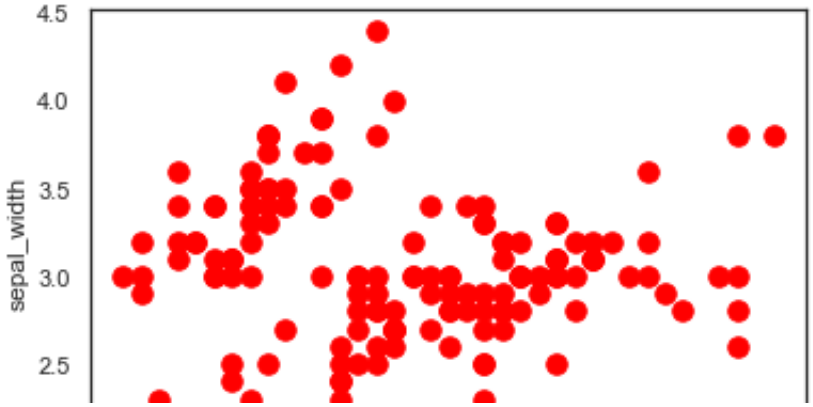
```
*c* argument looks like a single numeric RGB or RGBA sequence, which should be avoided as value
-mapping will have precedence in case its length matches with *x* & *y*.  Please use the *color
* keyword-argument or provide a 2-D array with a single row if you intend to specify the same R
GB or RGBA value for all points.
```

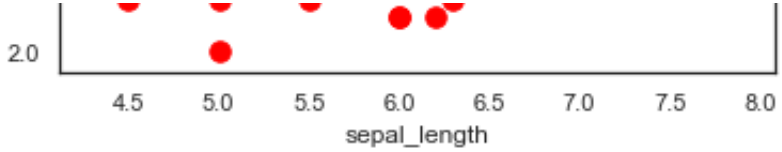




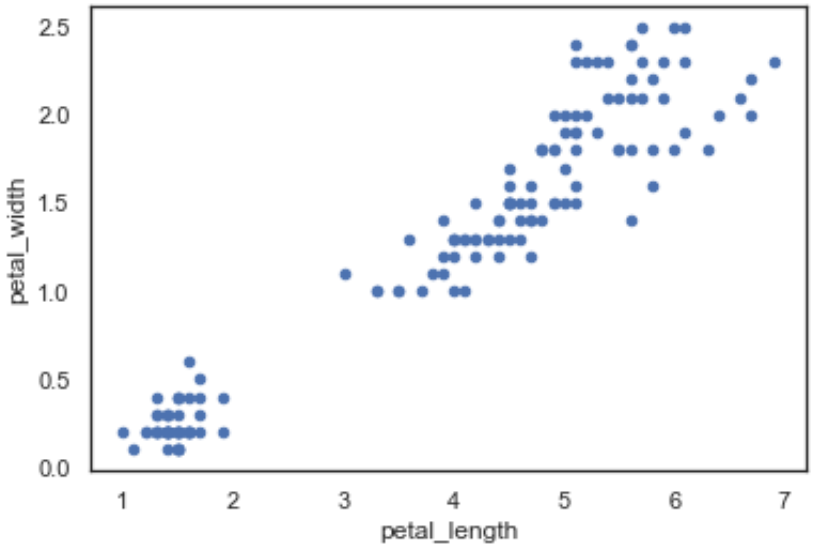

3
2
1
4.5
5.0
5.5
6.0
6.5
7.0
7.5
8.0
sepal_length

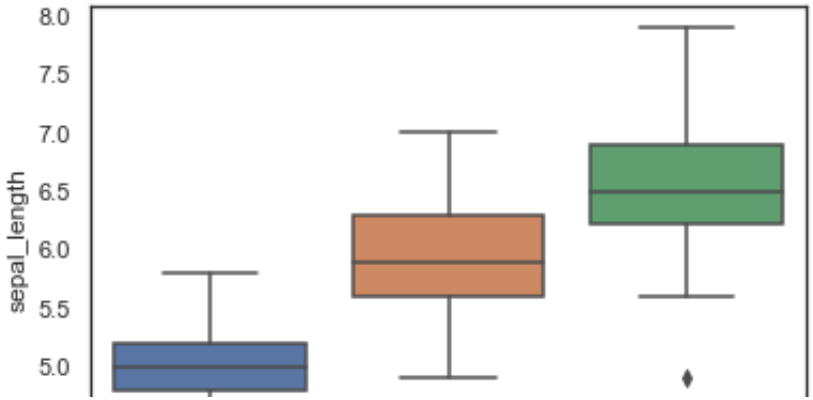

4.5
4.0
3.5
3.0
2.5
2.0
sepal_width
4.5
5.0
5.5
6.0
6.5
7.0
7.5
8.0
sepal_length

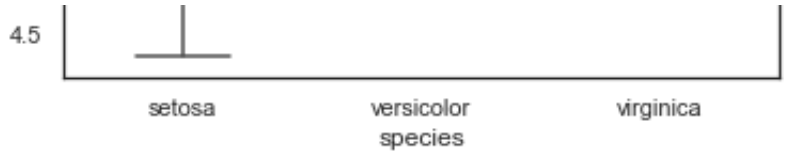

4.5
4.0
3.5
3.0
2.5
sepal_width

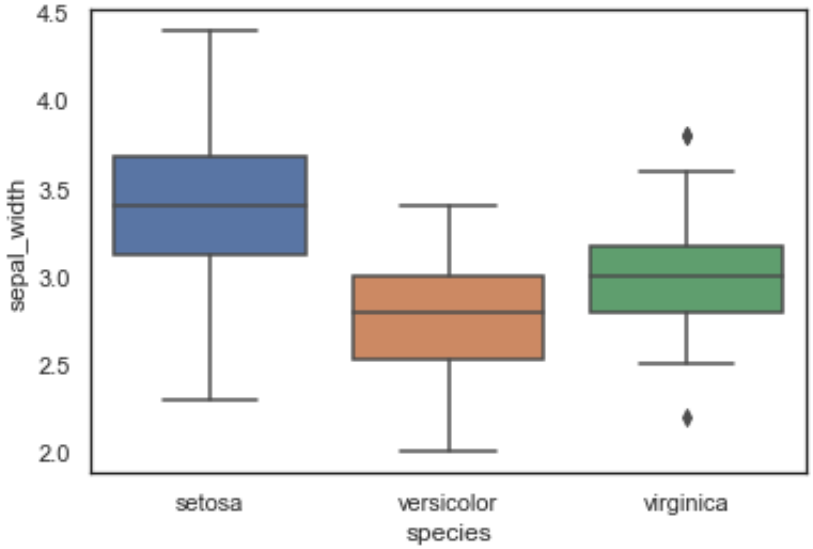


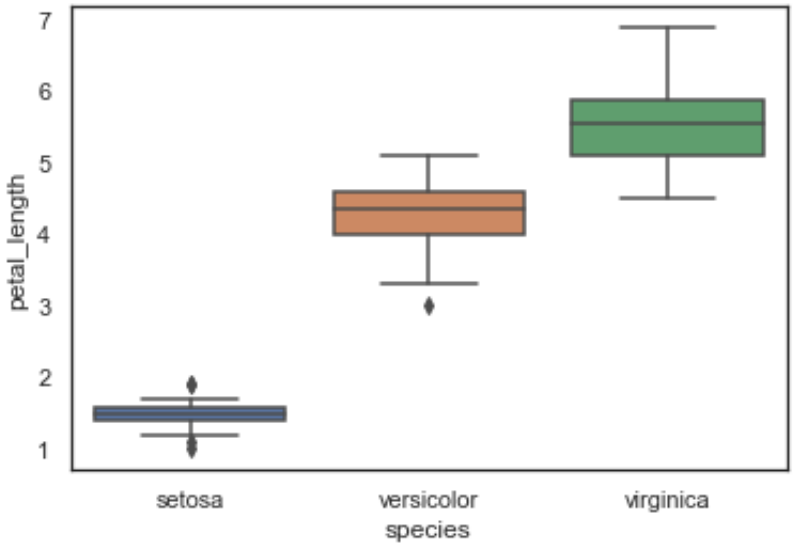

```
*c* argument looks like a single numeric RGB or RGBA sequence, which should be avoided as value
-mapping will have precedence in case its length matches with *x* & *y*.  Please use the *color
* keyword-argument or provide a 2-D array with a single row if you intend to specify the same R
GB or RGBA value for all points.
```

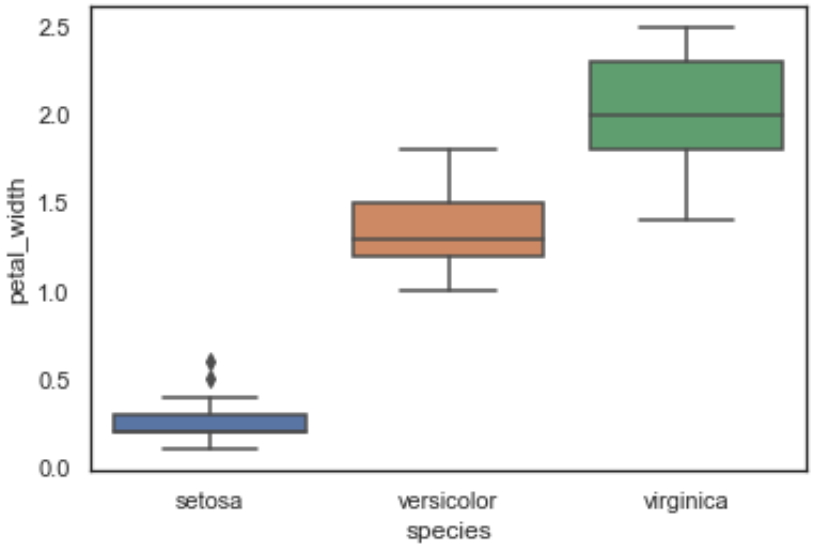





4.5
setosa
versicolor
virginica
species


4.5
4.0
3.5
3.0
2.5
2.0
sepal_width
setosa
versicolor
virginica
species


7
6
5
4
3
2
1
petal_length
setosa
versicolor
virginica
species



Summary

In [61]:
```
# According to our plot, the medians are vary
# and we can see that the sepal is are longer
# than the petal length in according to
# sepal_length_width and petal_length_width.
```