

```
import pandas as pd
iris = pd.read_csv('https://gist.githubusercontent.com/curran/a08a1080b88344b0c8
iris.head()
```

	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

```
iris.shape
```

```
(150, 5)
```

```
iris.describe()
```

	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

```
iris.info()
```

```
<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
```

```
# Checking Missing Values
```

```
iris.isnull().sum()
```

```
sepal_length    0
sepal_width     0
petal_length    0
petal_width     0
species         0
dtype: int64
```

```
# Checking duplicates
```

```
iris_duplicates = iris.drop_duplicates(subset = "species",)
iris_duplicates
```

	sepal_length	sepal_width	petal_length	petal_width	species
0	5.1	3.5	1.4	0.2	setosa
50	7.0	3.2	4.7	1.4	versicolor
100	6.3	3.3	6.0	2.5	virginica

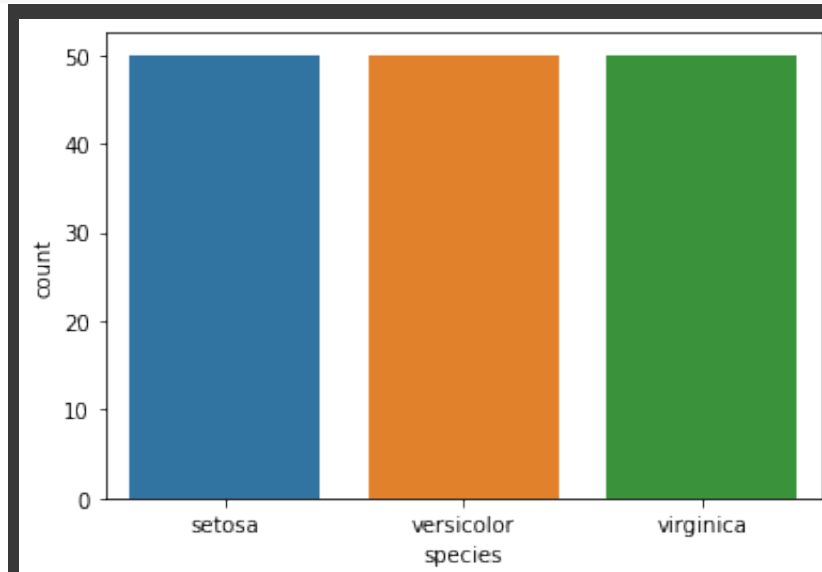
```
# Chekc balance
```

```
iris.value_counts("species")
```

```
species
setosa      50
versicolor  50
virginica   50
dtype: int64
```

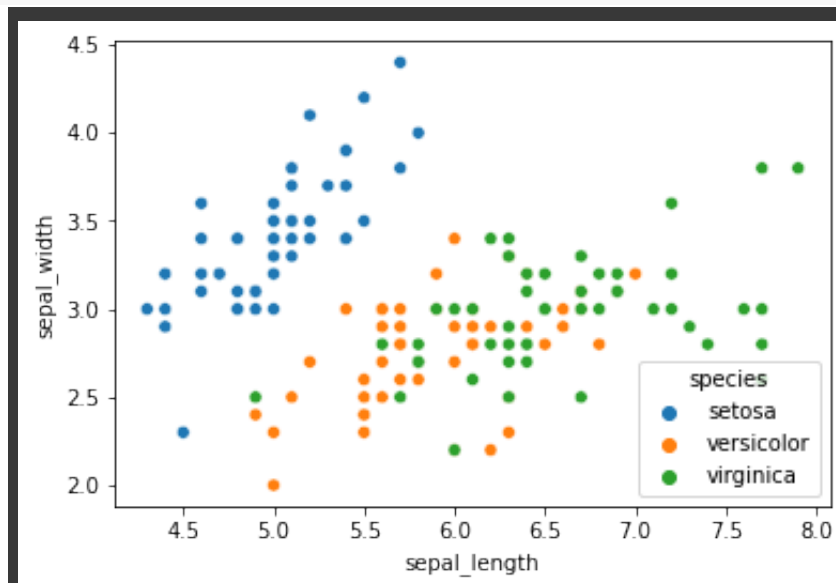
```
import seaborn as sns
import matplotlib.pyplot as plt

sns.countplot(x = 'species', data=iris)
plt.show()
```



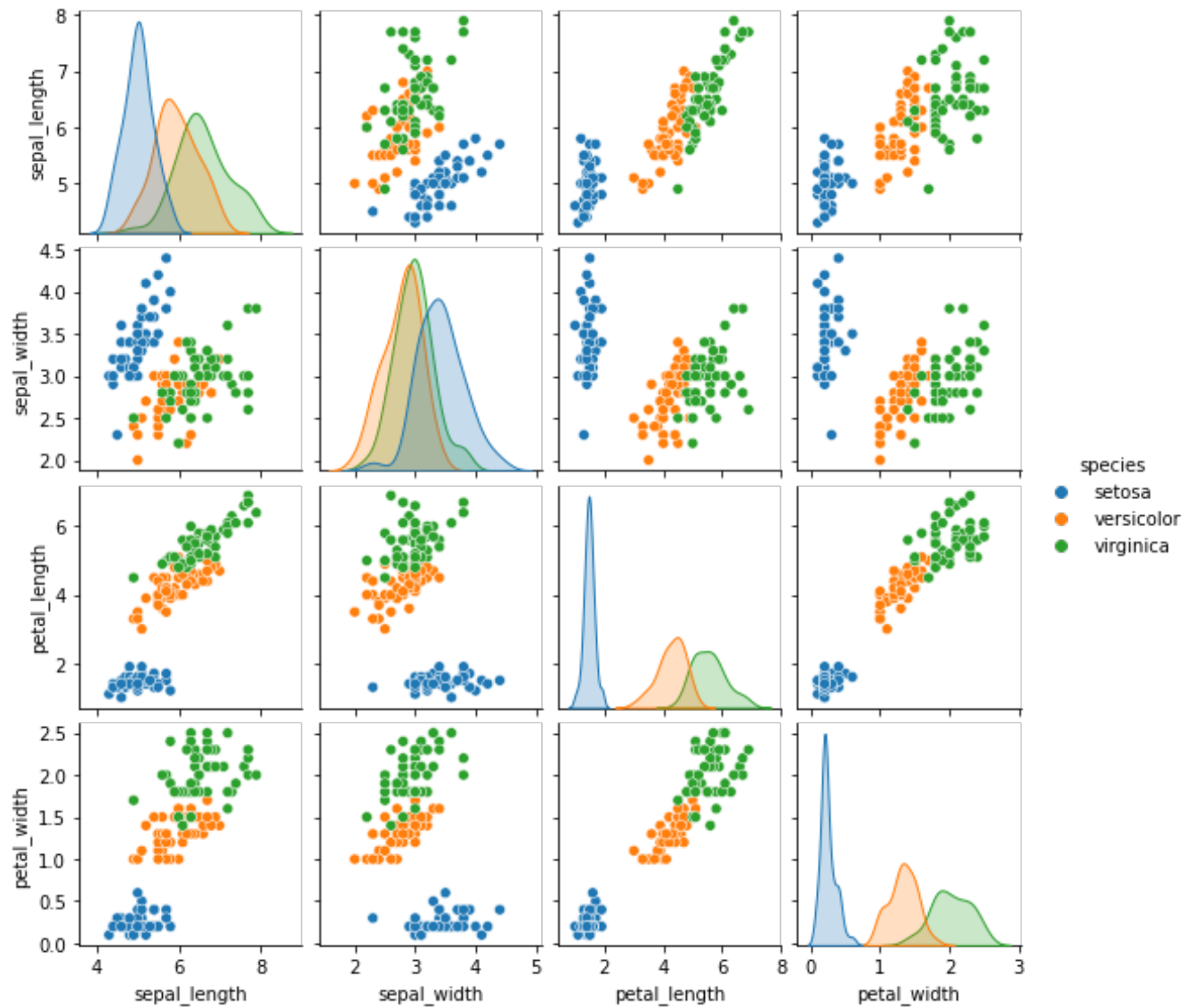
```
from matplotlib.transforms import BboxTransform
# Scatterplot the length and with of sepal.
import seaborn as sns
import matplotlib.pyplot as plt

sns.scatterplot(x='sepal_length', y='sepal_width', hue="species", data=iris, )
plt.show()
```



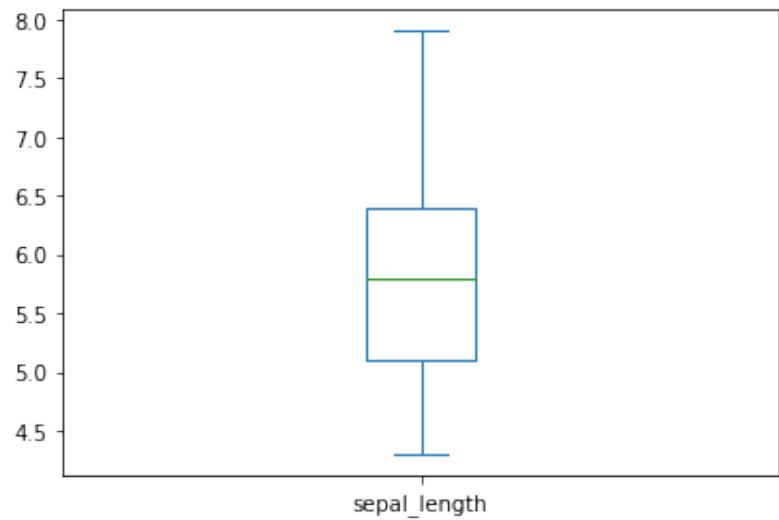
```
# Relationship between the species on the petal length, petal width, sepal length  
sns.pairplot(iris, aspect = 1,  
             hue='species', height=2)
```

<seaborn.axisgrid.PairGrid at 0x7f78cab7a310>



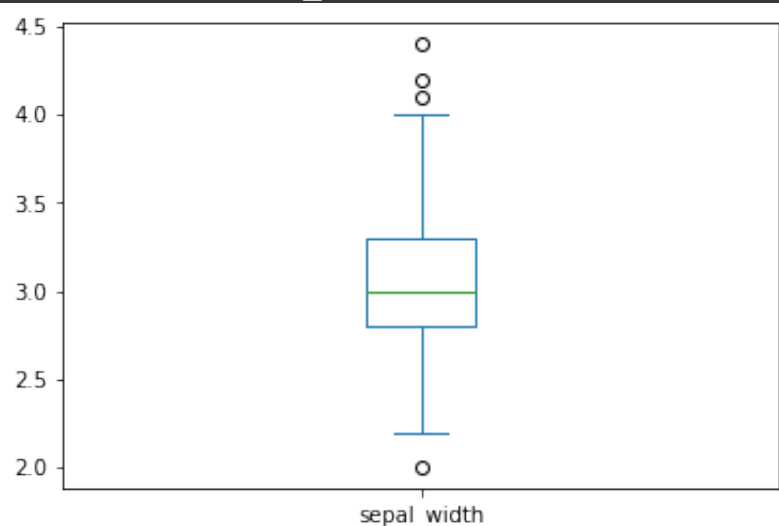
```
# Boxplot for seeing median, quantilez and max and min values  
iris['sepal_length'].plot(kind='box')
```

<matplotlib.axes._subplots.AxesSubplot at 0x7f78c5c45990>



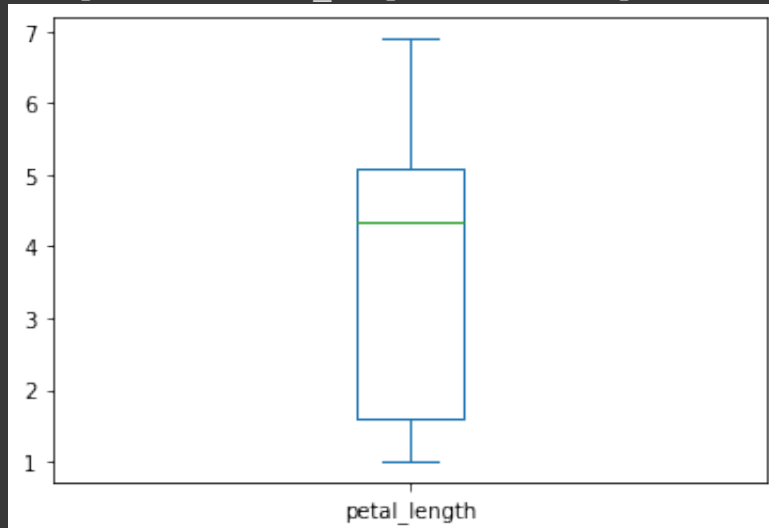
```
iris['sepal_width'].plot(kind='box')
```

<matplotlib.axes._subplots.AxesSubplot at 0x7f78c5bdefd0>



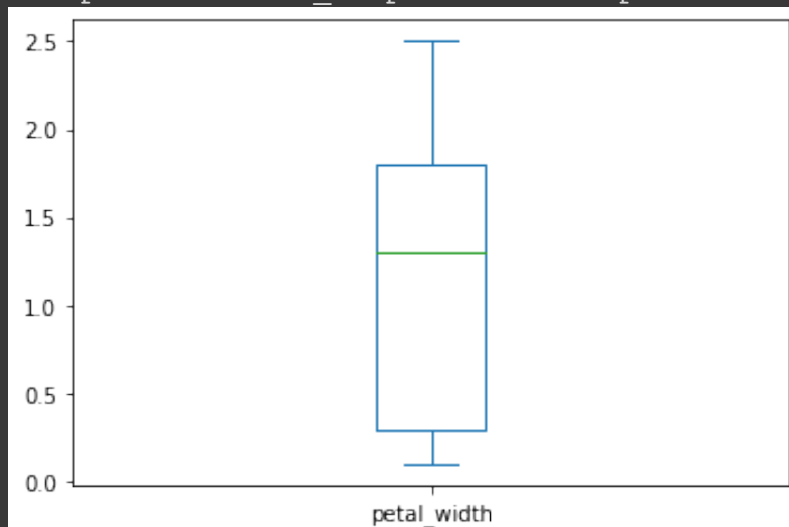
```
iris['petal_length'].plot(kind='box')
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f78c5b2a710>
```



```
iris['petal_width'].plot(kind='box')
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f78c5badf10>
```



```
# New feature
iris['petal_width_x_length'] = iris['petal_width'] * iris['petal_width']
iris['petal_width_x_length']
```

```
0      0.04
1      0.04
2      0.04
3      0.04
4      0.04
```

```
...
145    5.29
146    3.61
147    4.00
148    5.29
149    3.24
```

```
Name: petal_width_x_length, Length: 150, dtype: float64
```

[Colab paid products](#) - [Cancel contracts here](#)

