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

```
df=pd.read_csv("https://gist.githubusercontent.com/netj/8836201/raw/6f9306ad21398ea43cba4f
```

```
df.head()
```

	sepal.length	sepal.width	petal.length	petal.width	variety
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

```
df.tail()
```

	sepal.length	sepal.width	petal.length	petal.width	variety
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

```
df.shape
```

```
(150, 5)
```

```
df.columns
```

```
Index(['sepal.length', 'sepal.width', 'petal.length', 'petal.width',
       'variety'],
      dtype='object')
```

```
df["sepal.length"].unique()
```

```
array([5.1, 4.9, 4.7, 4.6, 5. , 5.4, 4.4, 4.8, 4.3, 5.8, 5.7, 5.2, 5.5,
       4.5, 5.3, 7. , 6.4, 6.9, 6.5, 6.3, 6.6, 5.9, 6. , 6.1, 5.6, 6.7,
       6.2, 6.8, 7.1, 7.6, 7.3, 7.2, 7.7, 7.4, 7.9])
```

```
df["sepal.width"].unique()
```

```
array([3.5, 3. , 3.2, 3.1, 3.6, 3.9, 3.4, 2.9, 3.7, 4. , 4.4, 3.8, 3.3,
       4.1, 4.2, 2.3, 2.8, 2.4, 2.7, 2. , 2.2, 2.5, 2.6])
```

```
df["petal.length"].unique()
```

```
array([1.4, 1.3, 1.5, 1.7, 1.6, 1.1, 1.2, 1. , 1.9, 4.7, 4.5, 4.9, 4. ,
       4.6, 3.3, 3.9, 3.5, 4.2, 3.6, 4.4, 4.1, 4.8, 4.3, 5. , 3.8, 3.7,
       5.1, 3. , 6. , 5.9, 5.6, 5.8, 6.6, 6.3, 6.1, 5.3, 5.5, 6.7, 6.9,
       5.7, 6.4, 5.4, 5.2])
```

```
df["petal.width"].unique()
```

```
array([0.2, 0.4, 0.3, 0.1, 0.5, 0.6, 1.4, 1.5, 1.3, 1.6, 1. , 1.1, 1.8,
       1.2, 1.7, 2.5, 1.9, 2.1, 2.2, 2. , 2.4, 2.3])
```

```
df["sepal.length"].value_counts().sum()
```

```
150
```

```
df["sepal.length"].value_counts()
```

```
5.0    10
5.1     9
6.3     9
5.7     8
6.7     8
5.8     7
5.5     7
6.4     7
4.9     6
5.4     6
6.1     6
6.0     6
5.6     6
4.8     5
6.5     5
6.2     4
7.7     4
6.9     4
4.6     4
5.2     4
5.9     3
4.4     3
7.2     3
6.8     3
6.6     2
4.7     2
7.6     1
7.4     1
7.3     1
7.0     1
7.1     1
5.3     1
4.3     1
```

```
4.5     1
7.9     1
Name: sepal.length, dtype: int64
```

```
df["sepal.width"].value_counts()
```

```
3.0    26
2.8    14
3.2    13
3.4    12
3.1    11
2.9    10
2.7     9
2.5     8
3.5     6
3.3     6
3.8     6
2.6     5
2.3     4
3.6     4
3.7     3
2.4     3
2.2     3
3.9     2
4.4     1
4.0     1
4.1     1
4.2     1
2.0     1
Name: sepal.width, dtype: int64
```

```
df["petal.length"].value_counts()
```

```
1.4    13
1.5    13
5.1     8
4.5     8
1.6     7
1.3     7
5.6     6
4.7     5
4.9     5
4.0     5
4.2     4
5.0     4
4.4     4
4.8     4
1.7     4
3.9     3
4.6     3
5.7     3
4.1     3
5.5     3
6.1     3
5.8     3
3.3     2
5.4     2
6.7     2
```

```
5.3     2
5.9     2
6.0     2
1.2     2
4.3     2
1.9     2
3.5     2
5.2     2
3.0     1
1.1     1
3.7     1
3.8     1
6.6     1
6.3     1
1.0     1
6.9     1
3.6     1
6.4     1
Name: petal.length, dtype: int64
```

```
df["petal.width"].value_counts()
```

```
0.2    29
1.3    13
1.8    12
1.5    12
1.4     8
2.3     8
1.0     7
0.4     7
0.3     7
2.1     6
2.0     6
0.1     5
1.2     5
1.9     5
1.6     4
2.5     3
2.2     3
2.4     3
1.1     3
1.7     2
0.6     1
0.5     1
Name: petal.width, dtype: int64
```

```
df["sepal.length"].dtype
```

```
dtype('float64')
```

```
df["sepal.width"].dtype
```

```
dtype('float64')
```

```
df["petal.length"].dtype
```

```
dtype('float64')
```

```
df["petal.width"].dtype
```

```
dtype('float64')
```

PART 1

```
#LIST DOWN THE FEATURES AND THEIR TYPES
```

```
df.columns
```

```
Index(['sepal.length', 'sepal.width', 'petal.length', 'petal.width',
       'variety'],
      dtype='object')
```

```
df["sepal.length"].dtype
```

```
dtype('float64')
```

```
df["sepal.width"].dtype
```

```
dtype('float64')
```

```
df["petal.length"].dtype
```

```
dtype('float64')
```

```
df["petal.width"].dtype
```

```
dtype('float64')
```

```
sns.countplot(data=df,x="variety")
plt.title("Species Count")
```

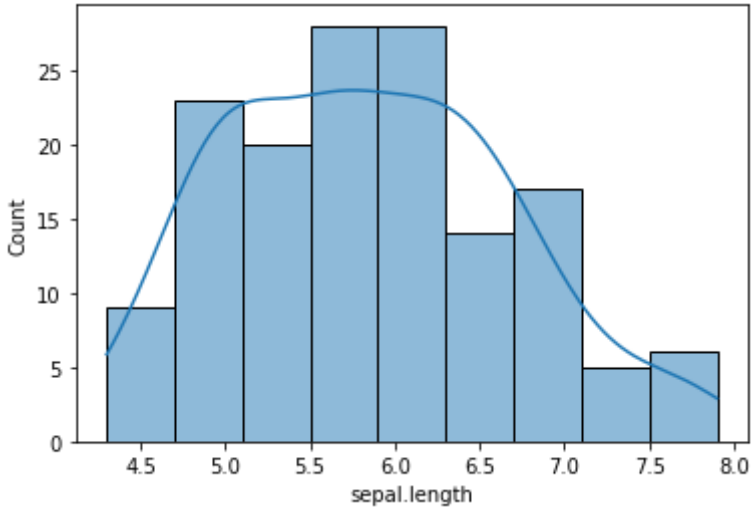
```
Text(0.5, 1.0, 'Species Count')
```

PART 2

```
#CREATE HISTOGRAM FOR EACH FEATURE IN DATASET TO ILLUSTRATE FEATRURE DISTRIBUTION
```

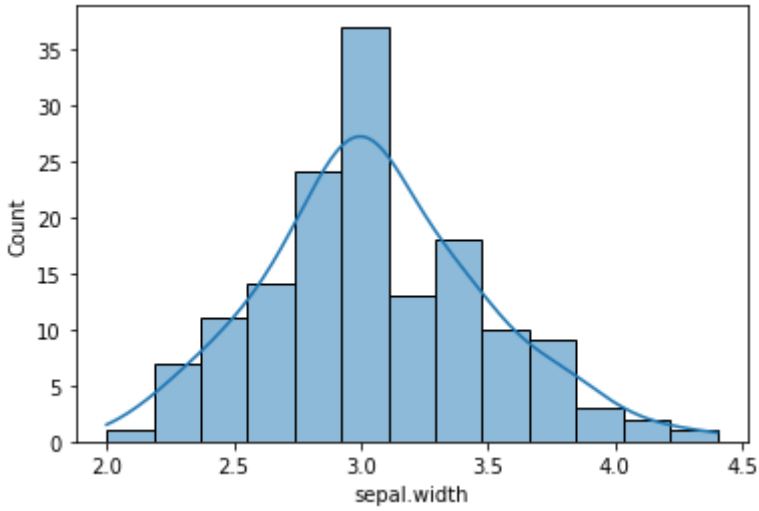
```
sns.histplot(data=df,x="sepal.length",kde=True)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7feb1a501310>
```



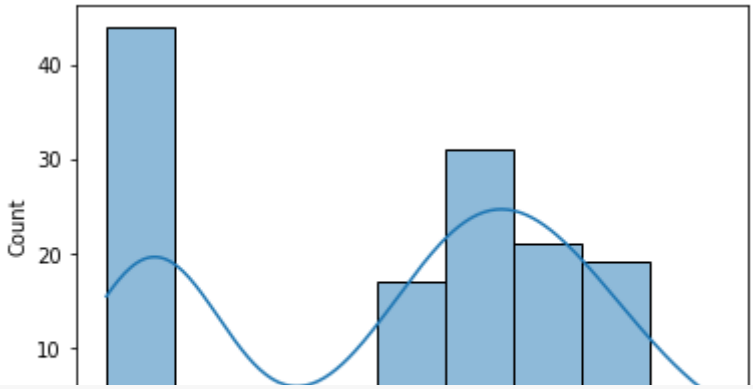

```
sns.histplot(data=df,x="sepal.width",kde=True)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7feb1a544690>
```



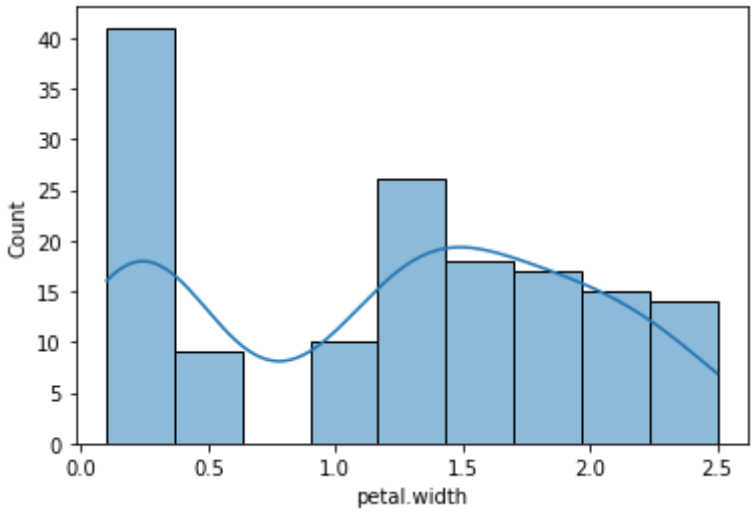

```
sns.histplot(data=df,x="petal.length",kde=True)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7feb1a3a7410>
```




```
sns.histplot(data=df,x="petal.width",kde=True)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7feb1a3d0a90>
```

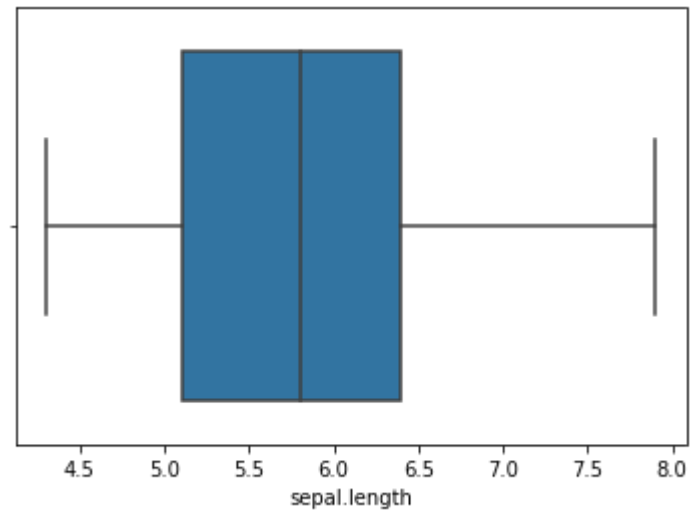


PART 3

```
#CREATE A BOX PLOT FOR EACH FEATURE IN THE DATASET
```

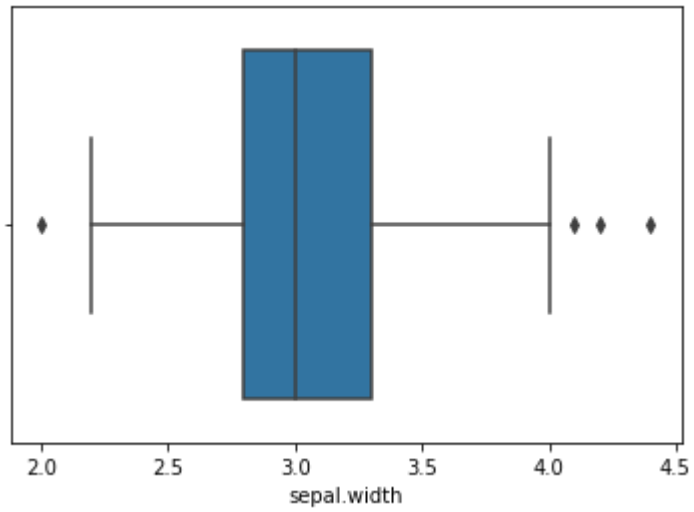
```
sns.boxplot(data=df,x="sepal.length")
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7feb1a2b2e90>
```



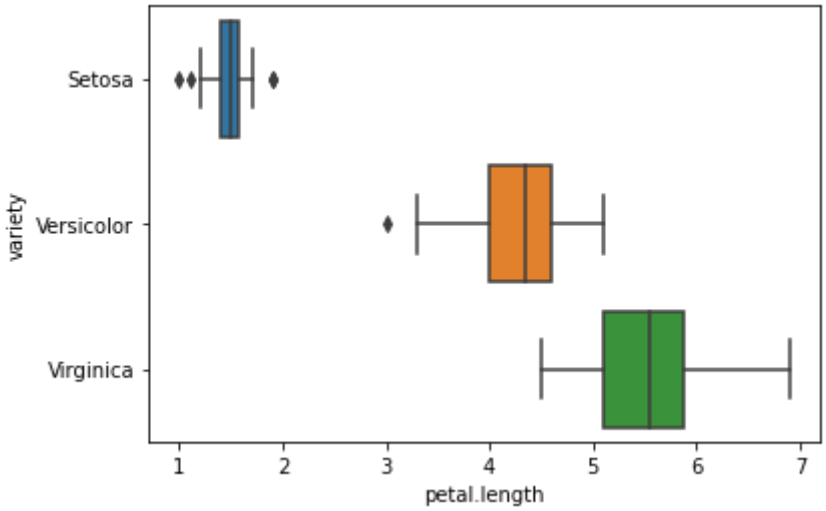
```
sns.boxplot(data=df,x="sepal.width")
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7feb1a7b7050>
```



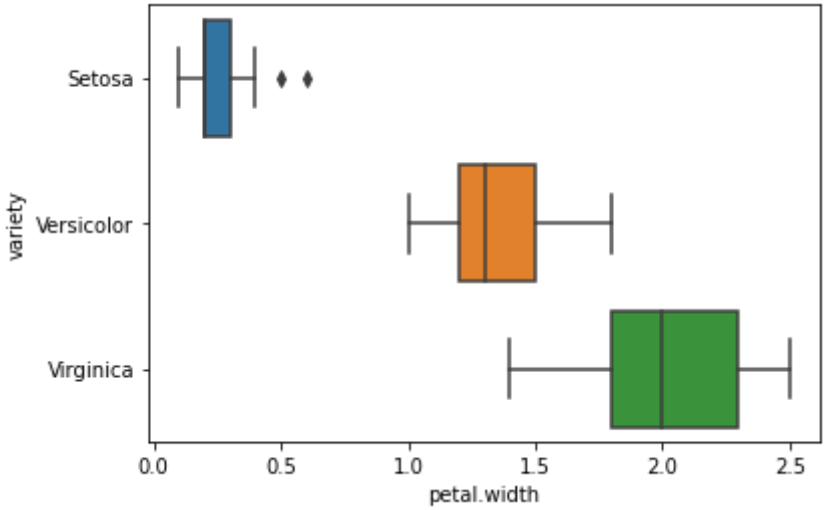

```
sns.boxplot(data=df,x="petal.length",y="variety")
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7feb1a6e6910>
```




```
sns.boxplot(data=df,x="petal.width",y="variety")
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7feb1a162ad0>
```

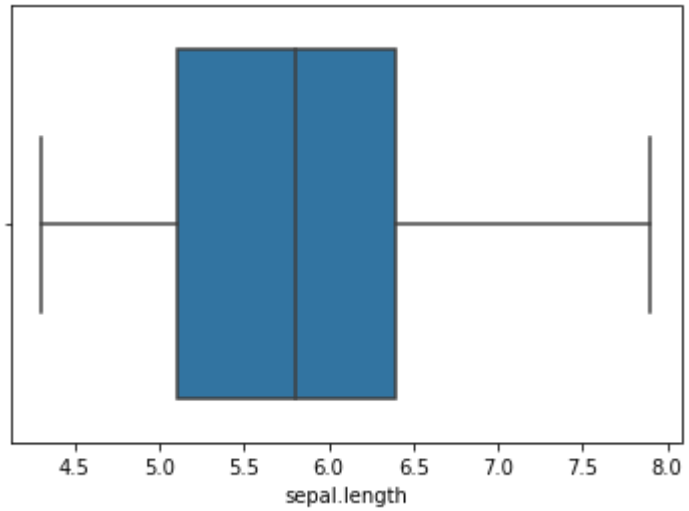


PART 4

```
#COMPARE DISTRIBUTIONS AND OUTLIERS
```

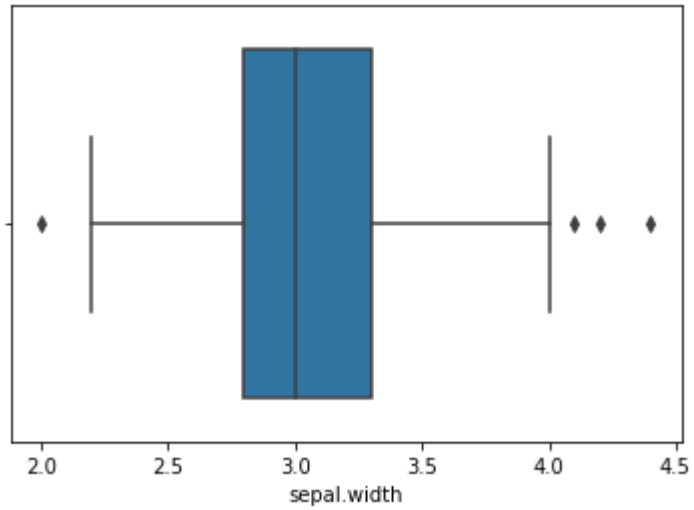
```
sns.boxplot(data=df,x="sepal.length")
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7feb1a10f690>
```



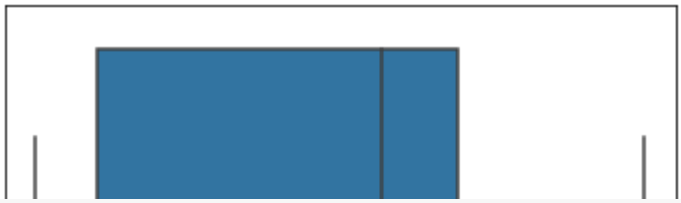

```
sns.boxplot(data=df,x="sepal.width")
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7feb1a0744d0>
```



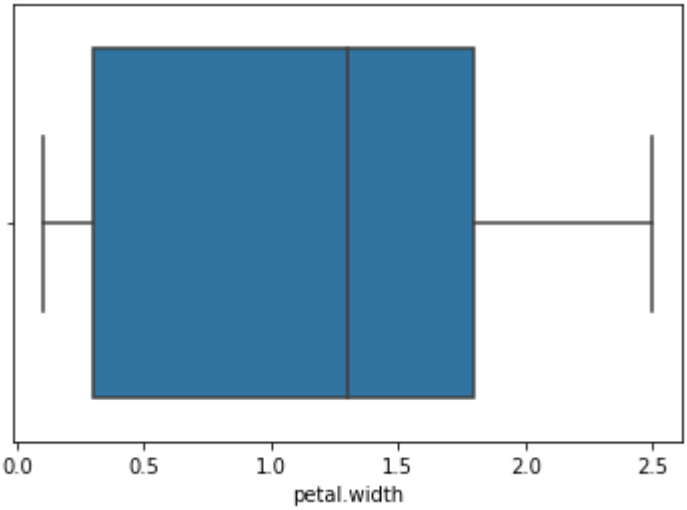

```
sns.boxplot(data=df,x="petal.length")
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7feb19fe0e50>
```



```
sns.boxplot(data=df,x="petal.width")
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7feb19fcf450>
```



✓ 0s completed at 10:03 PM