```
import pandas as pd
import seaborn as sns


# Load the Iris dataset
iris_df = pd.read_csv("https://raw.githubusercontent.com/uiuc-cse/data-fa14/gh-pages/data/iris.csv")



# Display the first few rows of the dataset to understand its structure
print(iris_df.head(50))
```

```
    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
```

```
# Check for missing values
print(iris_df.isnull().sum())
```

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

```
# Check data types and summary statistics
print(iris_df.info())
```

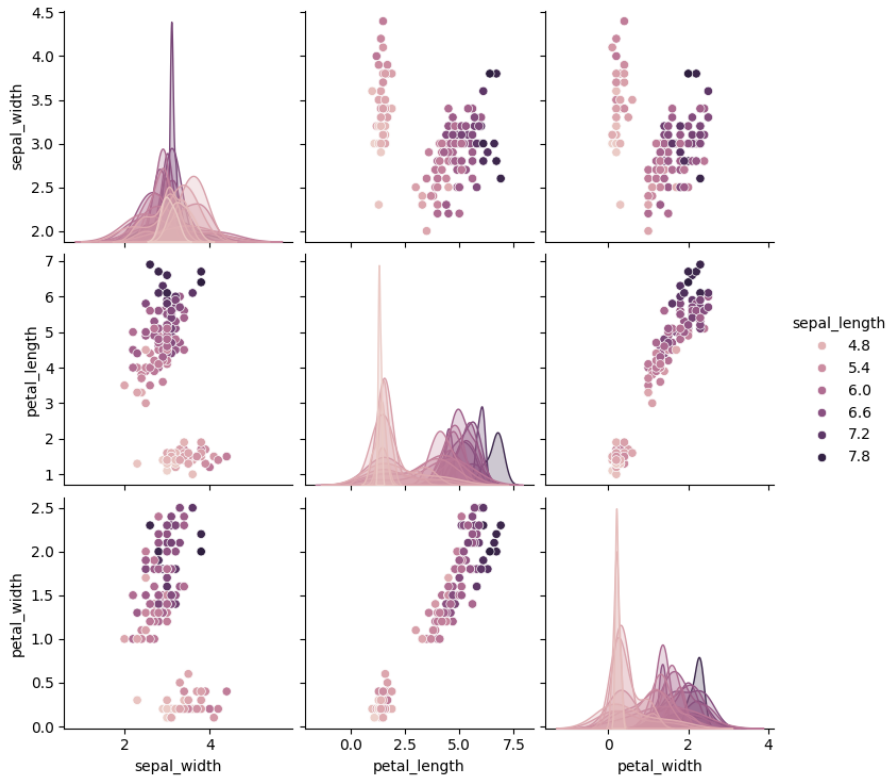
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 150 entries, 0 to 149
Data columns (total 5 columns):
 #   Column        Non-Null Count  Dtype  
---  ------        --------------  -----  
 0   sepal_length  150 non-null    float64
 1   sepal_width   150 non-null    float64
 2   petal_length  150 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
None
```

```
print(iris_df.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
```

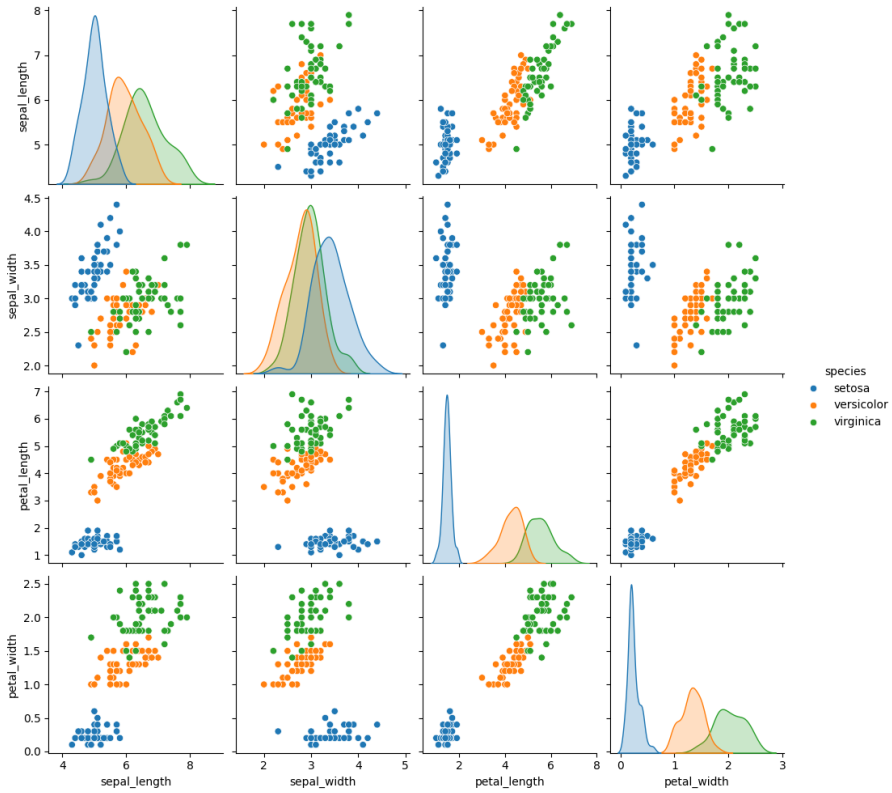
```
# Visualize by Seaborn
sns.pairplot(iris_df, hue="sepal_length")
```

```
<seaborn.axisgrid.PairGrid at 0x78c26baaf310>
```



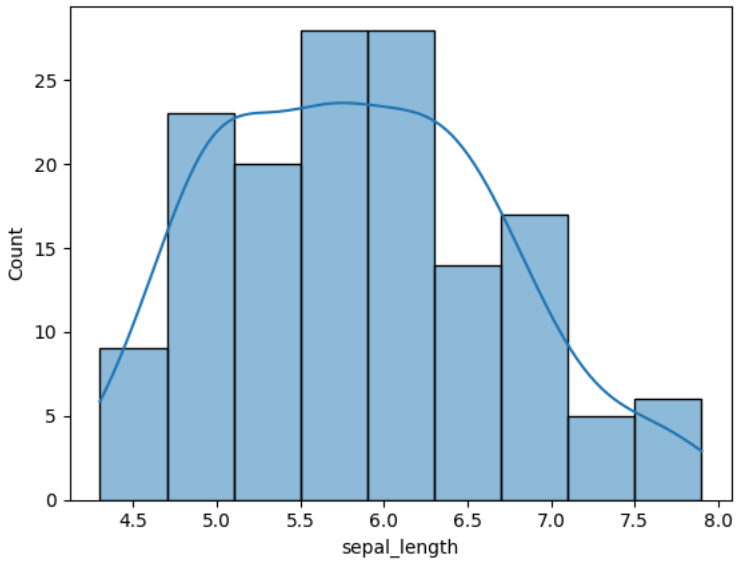

```
sns.pairplot(iris_df, hue="species")
```

```
<seaborn.axisgrid.PairGrid at 0x78c26cca7940>
```



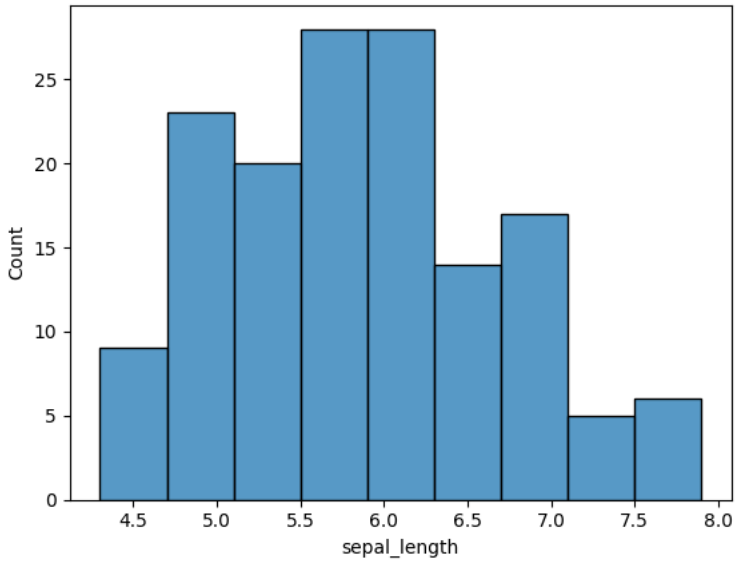

```
# Histogram and Density Plot for a single variable (e.g., Sepal Length)
sns.histplot(iris_df['sepal_length'], kde=True)
```

```
<Axes: xlabel='sepal_length', ylabel='Count'>
```



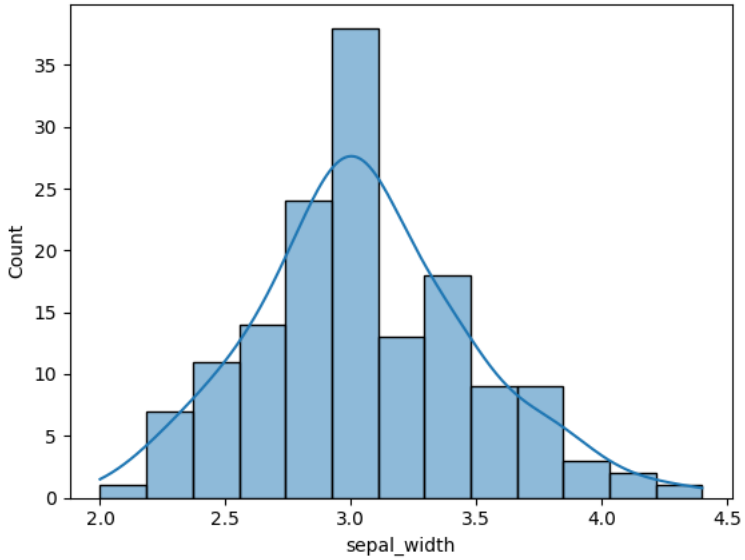

```
sns.histplot(iris_df['sepal_length'], kde=False)
```

```
<Axes: xlabel='sepal_length', ylabel='Count'>
```



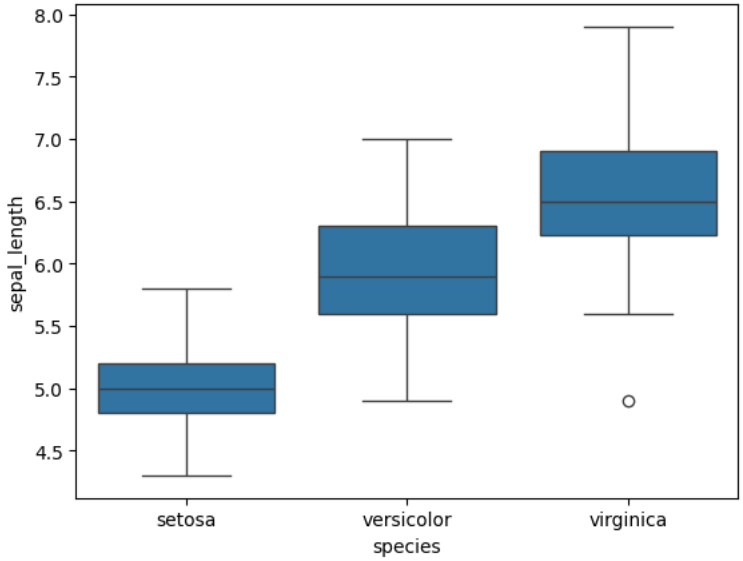

```
sns.histplot(iris_df['sepal_width'], kde=True)
```

```
<Axes: xlabel='sepal_width', ylabel='Count'>
```




```
# Boxplot of Sepal Length for each species
sns.boxplot(x='species', y='sepal_length', data=iris_df)
```

```
<Axes: xlabel='species', ylabel='sepal_length'>
```




```
# Group the data by species and calculate summary statistics for sepal length
# from pandas or without graph or without seaborn
summary_stats = iris_df.groupby('species')['sepal_length'].describe()

print(summary_stats)
```

```
              count   mean       std  min    25%  50%  75%  max
species
setosa         50.0  5.006  0.352490  4.3  4.800  5.0  5.2  5.8
versicolor     50.0  5.936  0.516171  4.9  5.600  5.9  6.3  7.0
virginica      50.0  6.588  0.635880  4.9  6.225  6.5  6.9  7.9
```

```
#  correlation matrix
correlation_matrix = iris_df.corr()

#  heatmap
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm')
```

```
<ipython-input-35-4bc4ba0a8be9>:2: FutureWarning: The default value of numeric_o
  correlation_matrix = iris_df.corr()
<Axes: >
```