

Data Visualization - iris

```
In [1]: import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

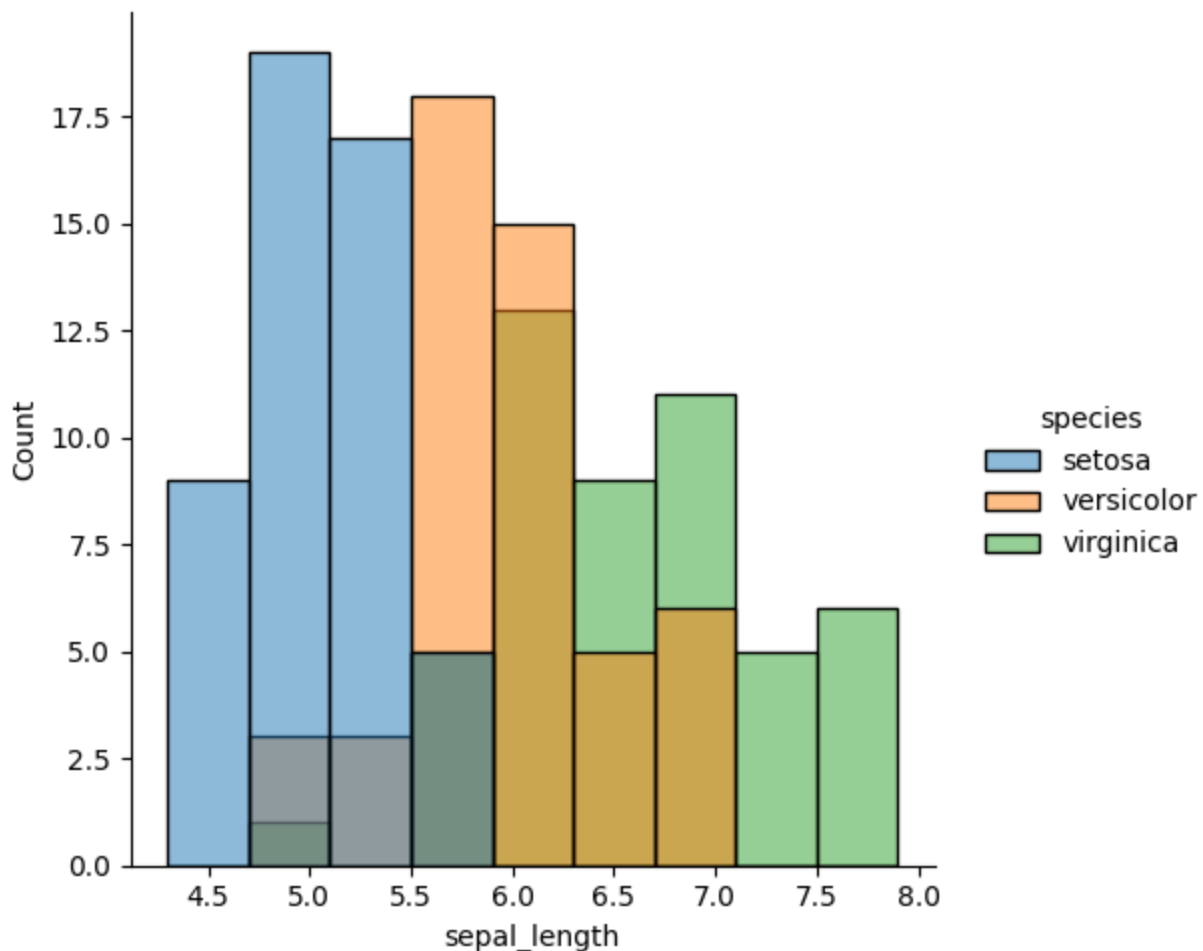
```
In [2]: df = sns.load_dataset('iris')
```

```
In [3]: df.columns
```

```
Out[3]: Index(['sepal_length', 'sepal_width', 'petal_length', 'petal_width',
              'species'],
              dtype='object')
```

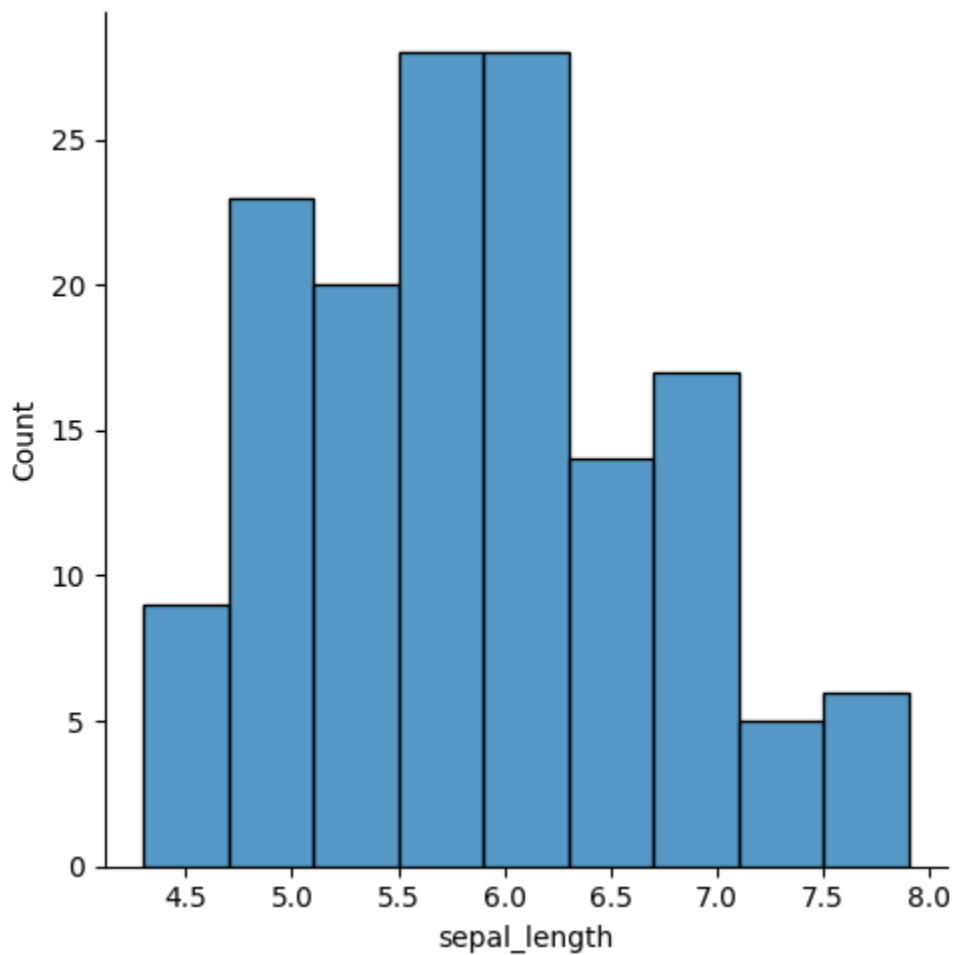
```
In [4]: sns.displot(x='sepal_length', data=df, hue='species')
```

```
Out[4]: <seaborn.axisgrid.FacetGrid at 0x263a2d716c0>
```



```
In [20]: sns.displot(x='sepal_length', data=df)
```

```
Out[20]: <seaborn.axisgrid.FacetGrid at 0x263ac6e38e0>
```

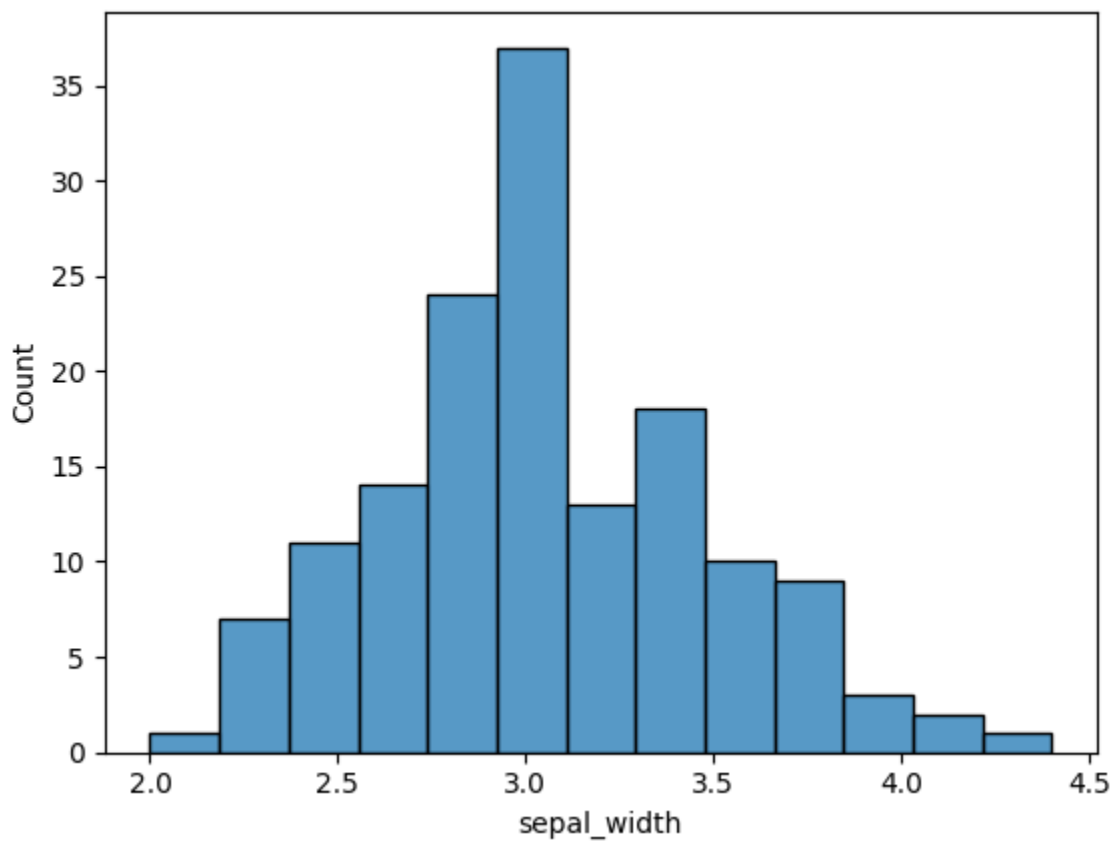


Inference:

- In the histogram for sepal length, we observe that the distribution is somewhat normal with a peak around 5.8 cm.
- Most of the iris flowers have a sepal length ranging from approximately 4.5 cm to 7.0 cm.
- There are relatively fewer flowers with extremely short or long sepal lengths, as evidenced by the lower frequencies at the tails of the distribution

```
In [19]: sns.histplot(x='sepal_width', data=df)
```

```
Out[19]: <AxesSubplot: xlabel='sepal_width', ylabel='Count'>
```

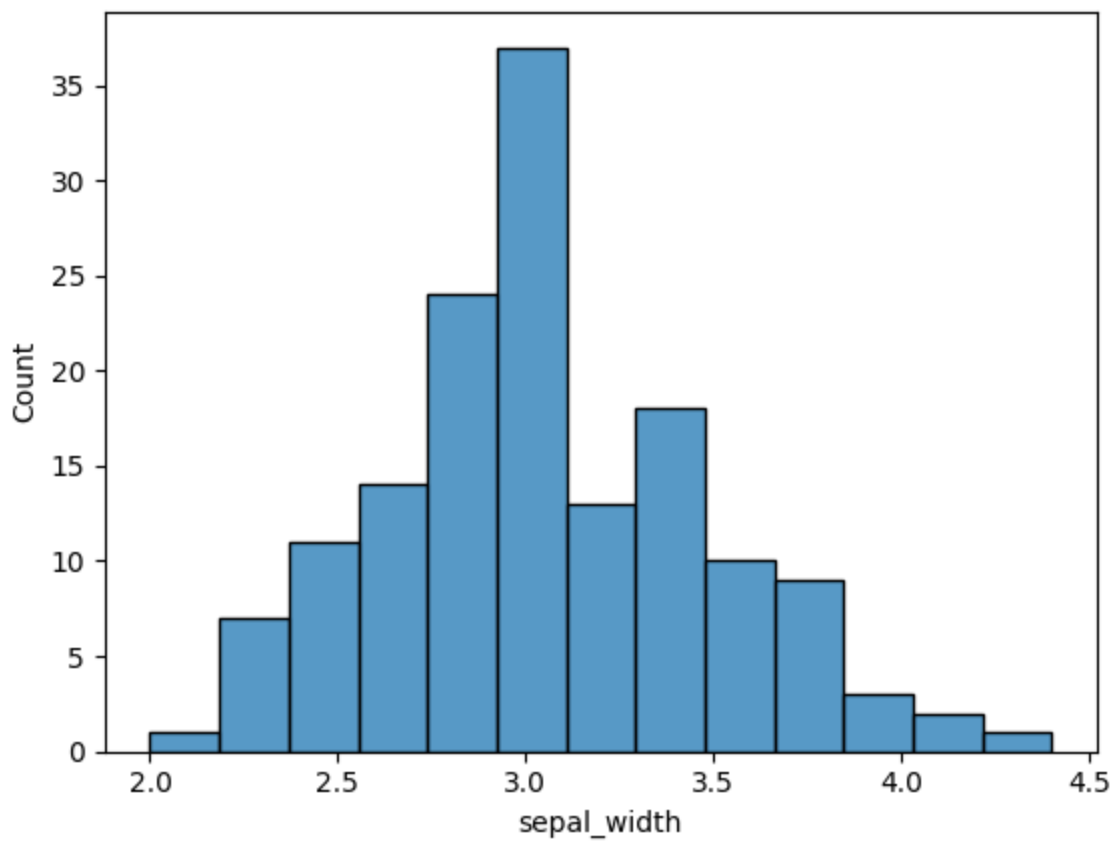


Inference:

- The histogram for sepal width shows a roughly normal distribution, albeit with some variation.
- The most common sepal width appears to be around 3.0 cm.
- There is a noticeable spread in sepal widths, ranging from approximately 2.0 cm to 4.5 cm. However, there's a slight skew towards higher sepal widths, as indicated by the slightly longer right tail of the distribution.

```
In [16]: sns.histplot(x='sepal_width', data=df)
```

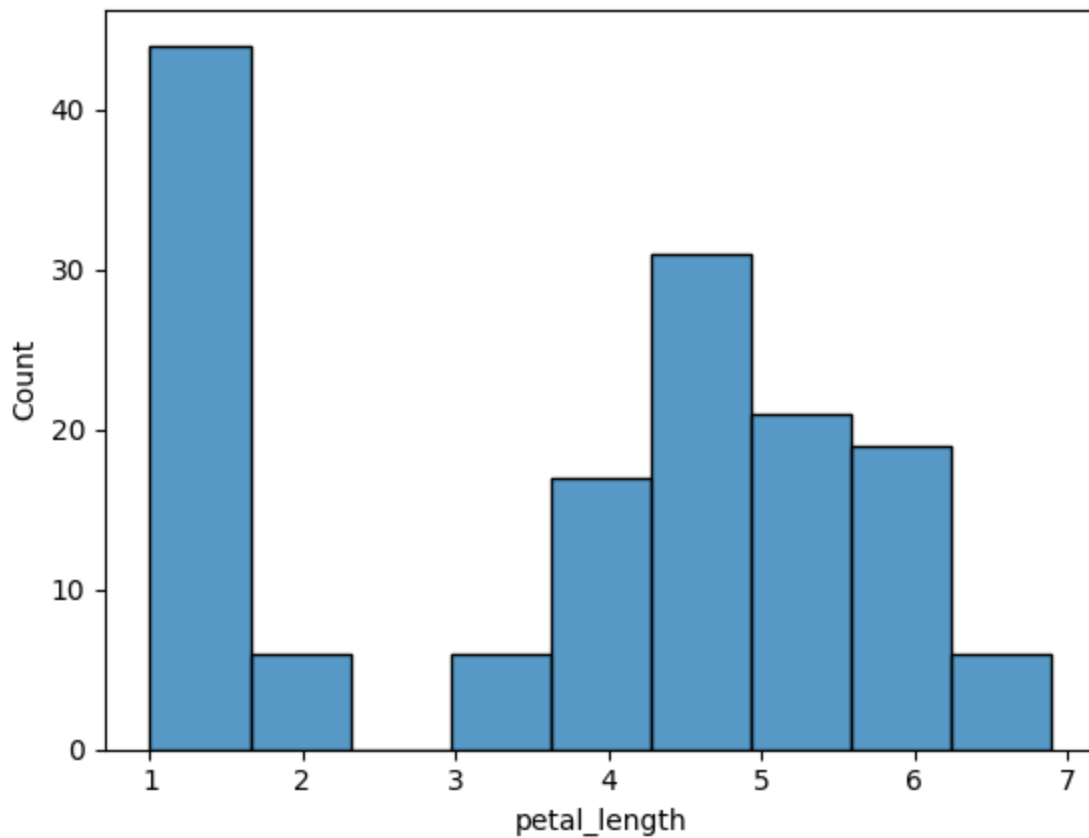
```
Out[16]: <AxesSubplot: xlabel='sepal_width', ylabel='Count'>
```



Inference: Shows the distribution of the sepal_width of the iris flowers. Count for sepal width 3.0 is highest. The has a little bit of skewness

```
In [8]: sns.histplot(x='petal_length', data=df)
```

```
Out[8]: <AxesSubplot: xlabel='petal_length', ylabel='Count'>
```

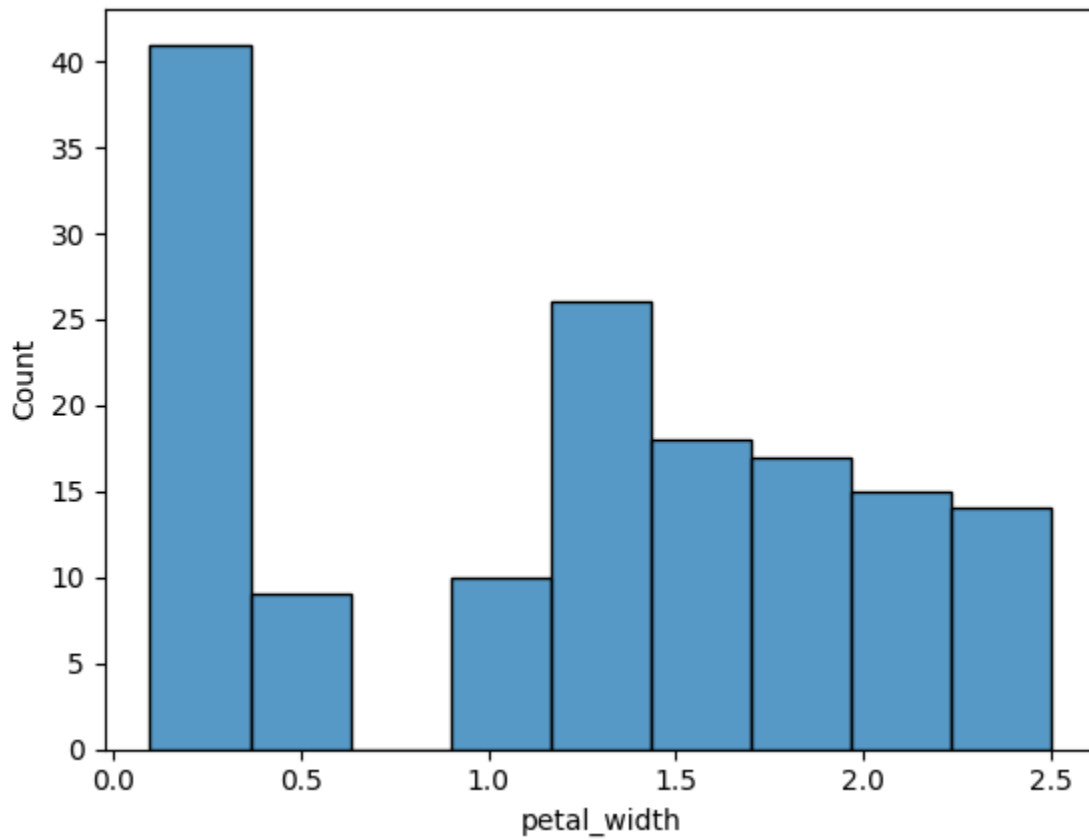


Inference:

- Data is not normally distributed
- Some of the petal length are ranging from 1-2
- Most data is in range of 3-7

```
In [17]: sns.histplot(x='petal_width', data=df)
```

```
Out[17]: <AxesSubplot: xlabel='petal_width', ylabel='Count'>
```



- Data is not normally distributed
- Some of the petal length are ranging from 0-0.5
- Most data is in range of 1.0-2.5

```
In [ ]:
```