In [2]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings('ignore')
```

Univariate -

Uni means one and variate means variable, so in univariate analysis, there is only one dependable variable. The objective of univariate analysis is to derive the data, define and summarize it, and analyze the pattern present in it. In a dataset, it explores each variable separately. It is possible for two kinds of variables- Categorical and Numerical.

In [5]:
```
# load dataset
df=sns.load_dataset('iris')
```

In [7]:
```
df.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

In [8]:
```
df.shape
```

```
(150, 5)
```

In [15]:
```
df['species'].value_counts()
```

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

Univariate Analysis

In [16]:
```
df_setosa=df.loc[df['species']=='setosa']
```

In [27]:
```
df_setosa.head(2)
```

	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

In [19]:
```
df_versicolor=df.loc[df['species']=='versicolor']
```

In [24]:
```
df_versicolor.head(2)
```

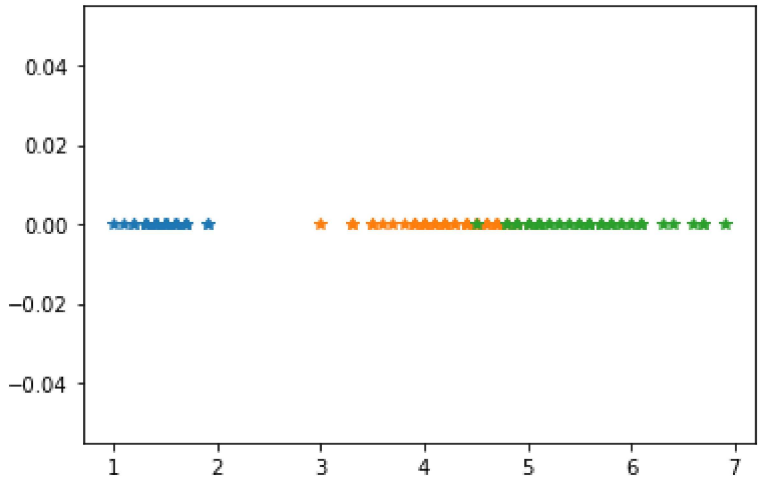
	sepal_length	sepal_width	petal_length	petal_width	species
50	7.0	3.2	4.7	1.4	versicolor
51	6.4	3.2	4.5	1.5	versicolor

In [22]:
```
df_virginica=df.loc[df['species']=='virginica']
```

In [26]:
```
df_virginica.head(2)
```

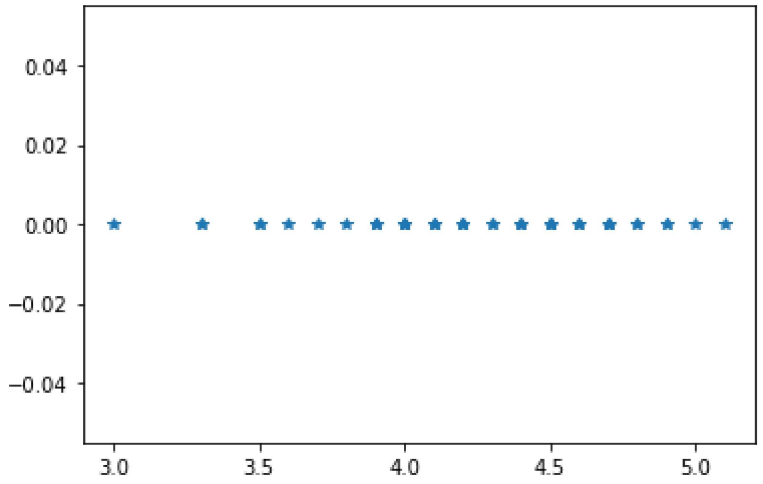
	sepal_length	sepal_width	petal_length	petal_width	species
100	6.3	3.3	6.0	2.5	virginica
101	5.8	2.7	5.1	1.9	virginica

In [33]:
```
plt.plot(df_setosa['petal_length'],np.zeros_like(df_setosa['petal_length']),'*')
plt.plot(df_versicolor['petal_length'],np.zeros_like(df_versicolor['petal_length']),'*')
plt.plot(df_virginica['petal_length'],np.zeros_like(df_virginica['petal_length']),'*')
plt.show()
```



In [34]:
```
plt.plot(df_versicolor['petal_length'],np.zeros_like(df_versicolor['petal_length']),'*')
```

```
[<matplotlib.lines.Line2D at 0x22c041f2b50>]
```



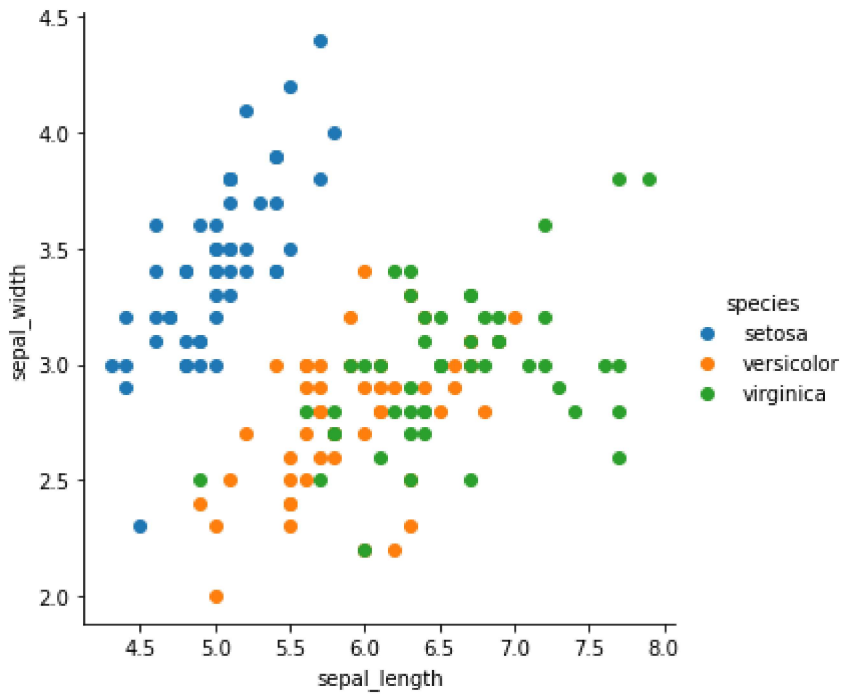
In [35]:
```
# if you see above there is some overlapping
# now we use Bivariate
```

Bivariate -

Bi means two and variate means variable, so here there are two variables. The analysis is related to cause and the relationship between the two variables.

Bivariate Analysis

In [39]:
```
sns.FacetGrid(df,hue='species',size=5).map(plt.scatter,'sepal_length','sepal_width').add_legend();
plt.show()
```



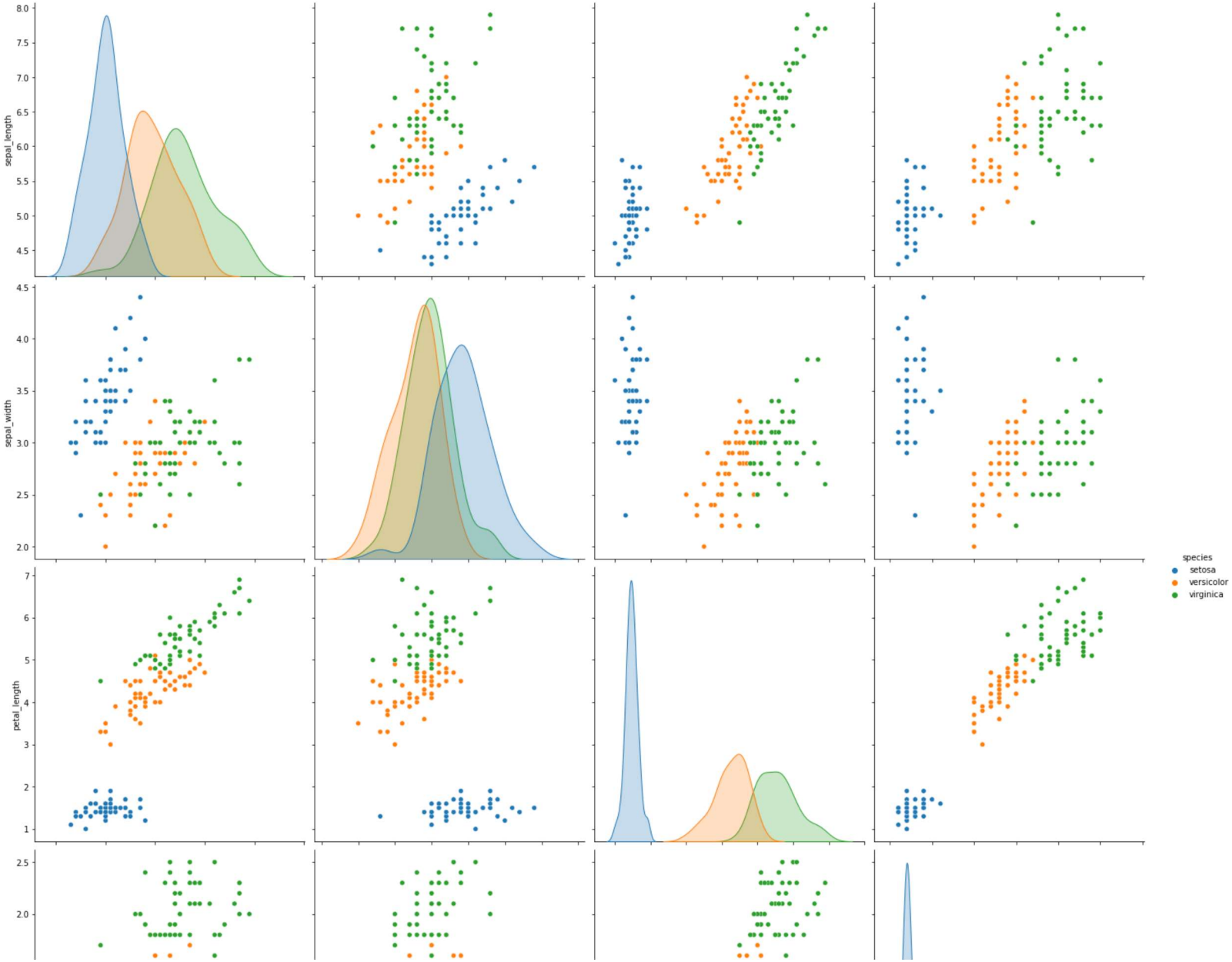
In [ ]:

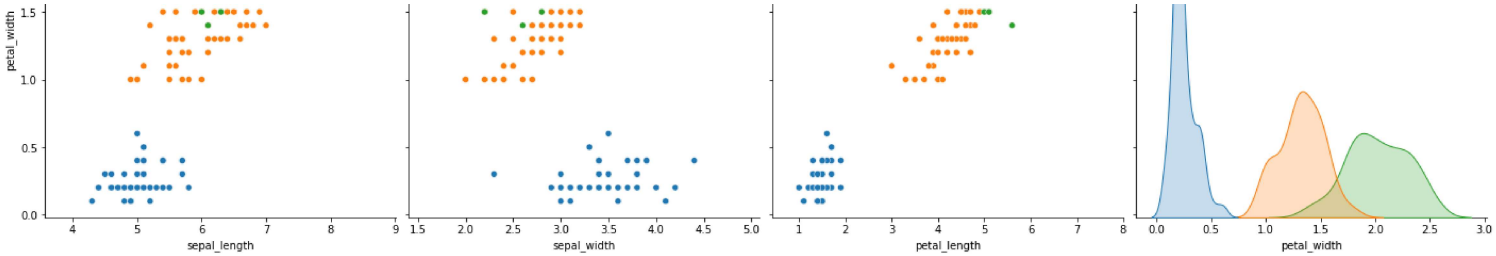
Multivariate -

Multivariate analysis is required when more than two variables have to be analyzed simultaneously. It is a tremendously hard task for the human brain to visualize a relationship among 4 variables in a graph and thus multivariate analysis is used to study more complex sets of data. Types of Multivariate Analysis include Cluster Analysis, Factor Analysis, Multiple Regression Analysis, Principal Component Analysis, etc. More than 20 different ways to perform multivariate analysis exist and which one to choose depends upon the type of data and the end goal to achieve.

In [42]:
```
sns.pairplot(df,hue='species',size=5)
```

```
<seaborn.axisgrid.PairGrid at 0x22c049e2d30>
```




petal_width
sepal_length
sepal_width
petal_length
petal_width

In [ ]:

In [ ]: