DATA HANDLING USING ‘Pandas’ and DATA VISUALIZATION USING ‘Seaborn’

Using the pandas function read_csv(), read the given ‘iris’ data set.

1. Use appropriate functions in pandas to display
   (i) Shape of the data set
   (ii) First 5 and last five rows of data set(head and tail)
   (iii) Size of dataset
   (iv) No:of samples available for each variety
   (v) Description of the data set(use describe

CODE

```
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import pandas as pd
col=['sepal_length','sepal_width','petal_length','petal_width','type']
iris=pd.read_csv("iris.csv",names=col)
print("First five rows")
print(iris.head())
print("**********")
print("Last five rows")
print(iris.tail())
print("**********")
print("shape:",iris.shape)
print("**********")
print("Size:",iris.size)
print("**********")
print("no of samples available for each type")
print(iris["type"].value_counts())
print("**********")
print(iris.describe())
```

OUTPUT

```
First five rows
   sepal_length  sepal_width  petal_length  petal_width     type
0  sepal.length  sepal.width  petal.length  petal.width  variety
1           5.1          3.5           1.4           .2   Setosa
2           4.9            3           1.4           .2   Setosa
3           4.7          3.2           1.3           .2   Setosa
4           4.6          3.1           1.5           .2   Setosa
*********
Last five rows
    sepal_length sepal_width petal_length petal_width       type
146          6.7           3          5.2         2.3  Virginica
147          6.3         2.5            5         1.9  Virginica
148          6.5           3          5.2           2  Virginica
149          6.2         3.4          5.4         2.3  Virginica
150          5.9           3          5.1         1.8  Virginica
*********
shape: (151, 5)
*********
Size: 755
*********
no of samples available for each type
Setosa        50
Versicolor    50
Virginica     50
variety        1
Name: type, dtype: int64
*********
       sepal_length sepal_width petal_length petal_width    type
count           151         151          151         151     151
unique           36          24           44          23       4
top               5           3          1.5          .2  Setosa
freq             10          26           13          29      50
```

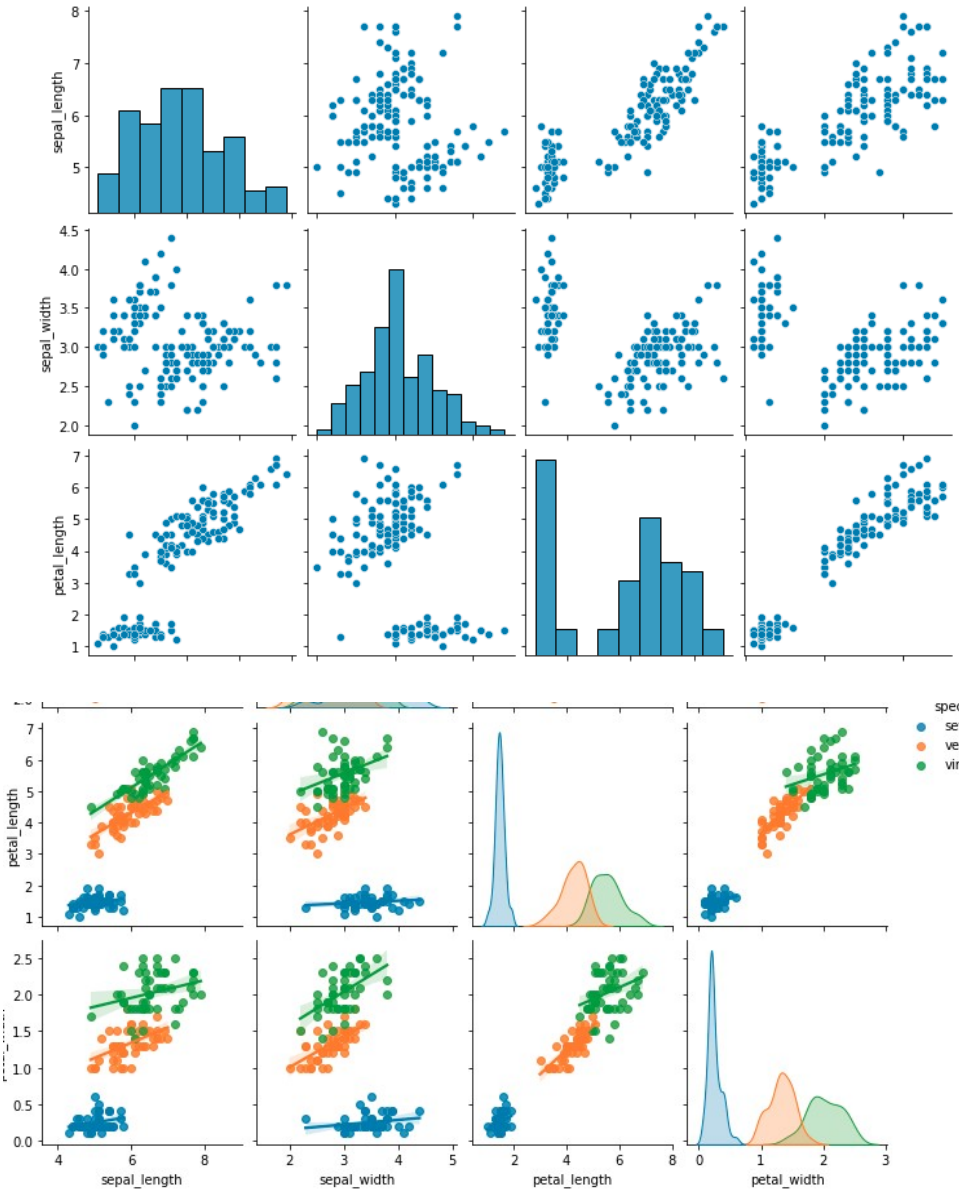
2. Use pairplot() function to display pairwise relationships between attributes. Try different kind of plots {***'scatter', 'kde', 'hist', 'reg'***} and different kind of markers

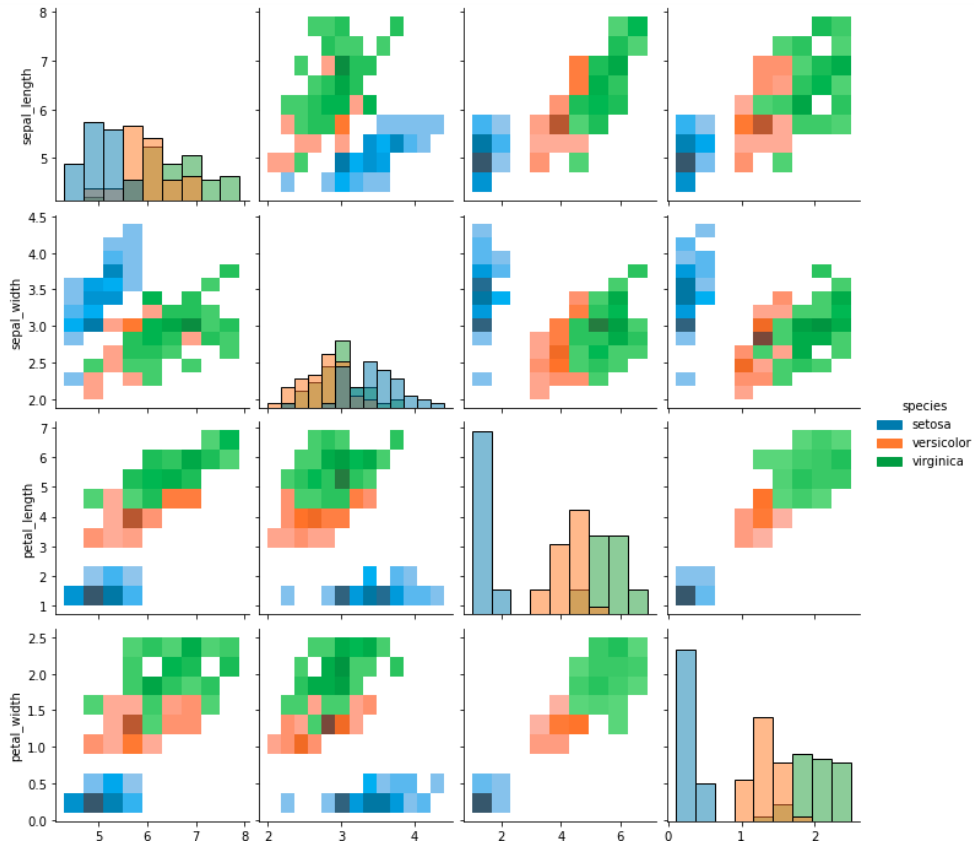
CODE

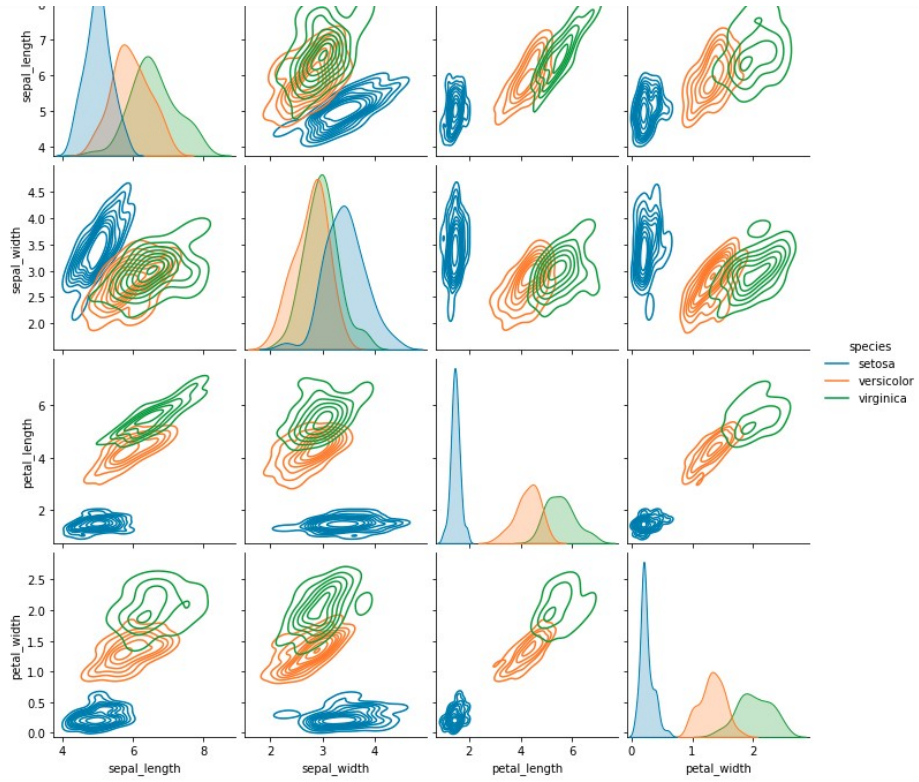
```
import matplotlib.pyplot as plt

%matplotlib inline



iris = sns.load_dataset('iris')

my_data_frame = pd.DataFrame(iris)



g = sns.pairplot(my_data_frame)
```

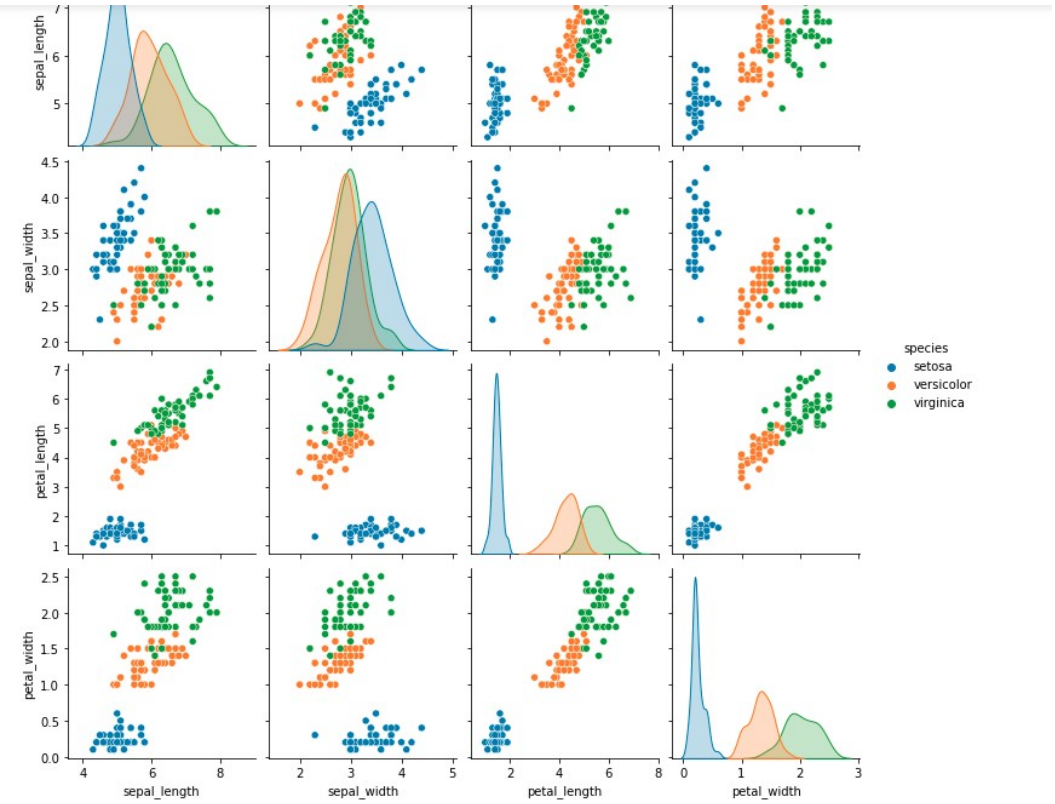
```
g = sns.pairplot(iris, kind="reg", hue="species")
g = sns.pairplot(iris, kind="hist", hue="species")
g = sns.pairplot(iris, kind="kde", hue="species")
g = sns.pairplot(iris, kind="scatter", hue="species")
```

OUTPUT




sepal_length
sepal_width
petal_length
petal_width
species
setosa
versicolor
virginica


sepal_length
sepal_width
petal_length
petal_width
species
setosa
versicolor
virginica



3. using the iris data set,get familiarize with functions:

1)displot()

2) histplot()

3) relplot()

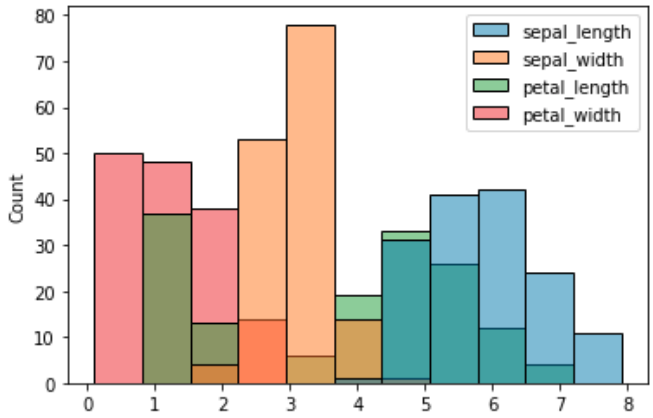
CODE

```
sns.displot(data=iris)
sns.histplot(data=iris)
sns.relplot(data=iris)
```

OUTPUT

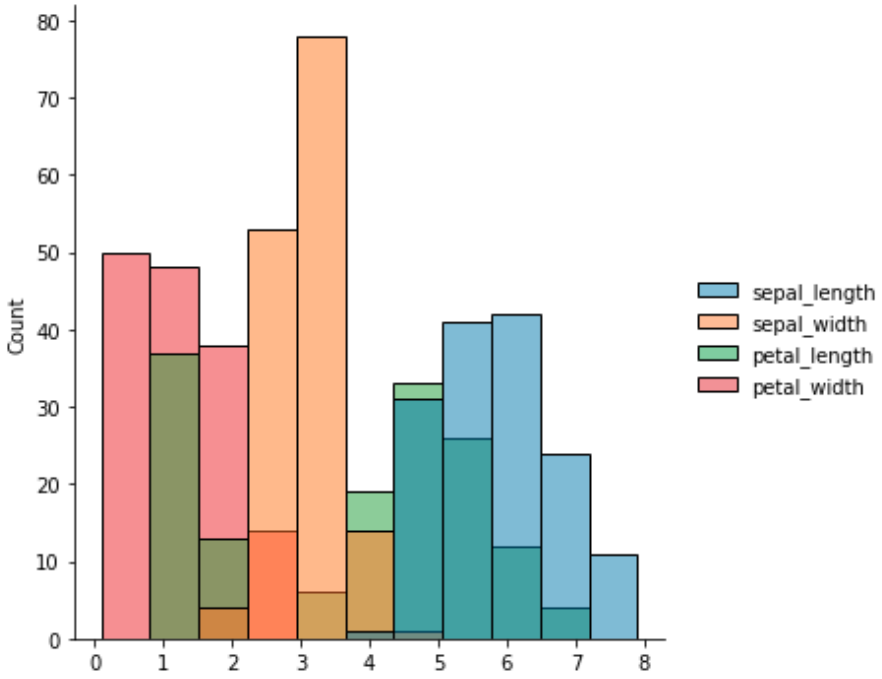
```
sns.histplot(data=iris)
```

```
<AxesSubplot:ylabel='Count'>
```



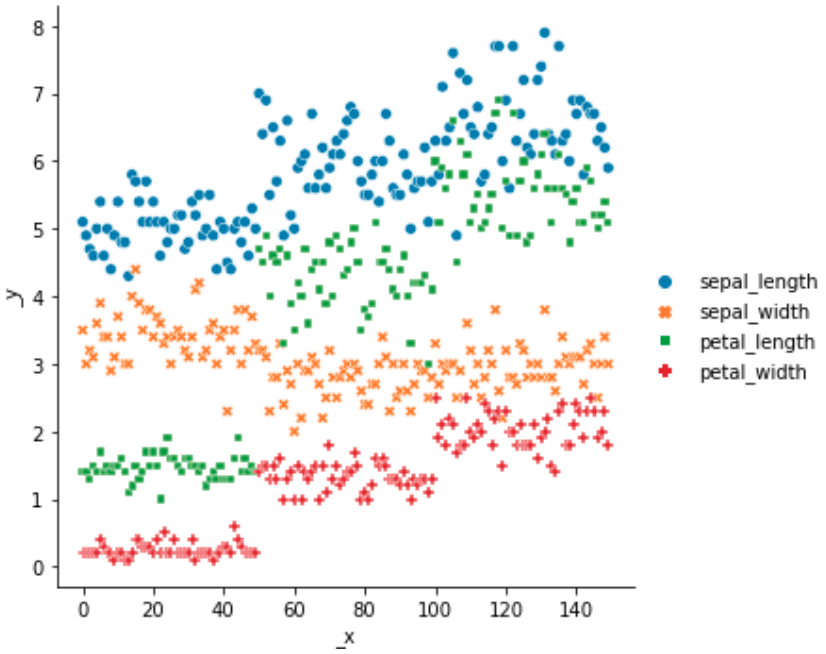

```
sns.displot(data=iris)
```

```
<seaborn.axisgrid.FacetGrid at 0x7fe3290bca00>
```



```
sns.relplot(data=iris)
```

```
<seaborn.axisgrid.FacetGrid at 0x7fe32d2cf130>
```



OR

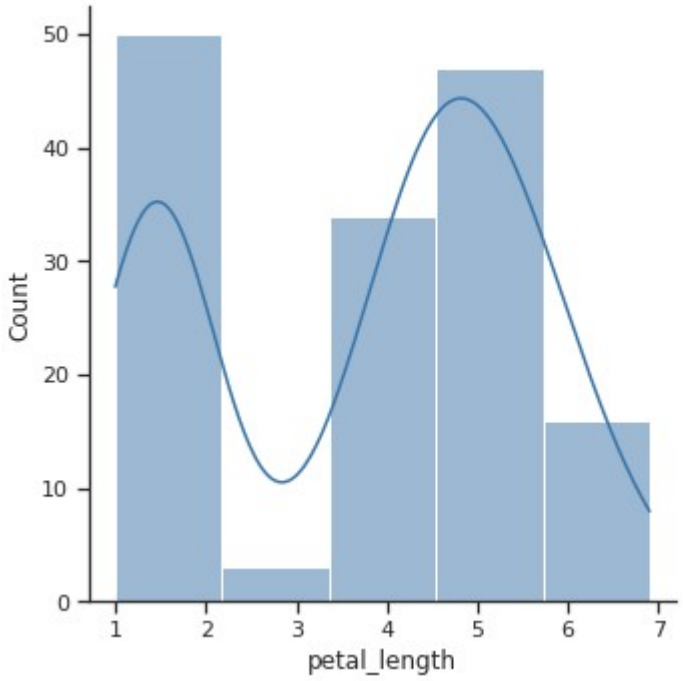
```
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import pandas as pd
col=['sepal_length','sepal_
df = sns.load_dataset('iris')
df.head()
sns.displot(x = 'petal_length',kde=True,bins = 5 , data =df)
```

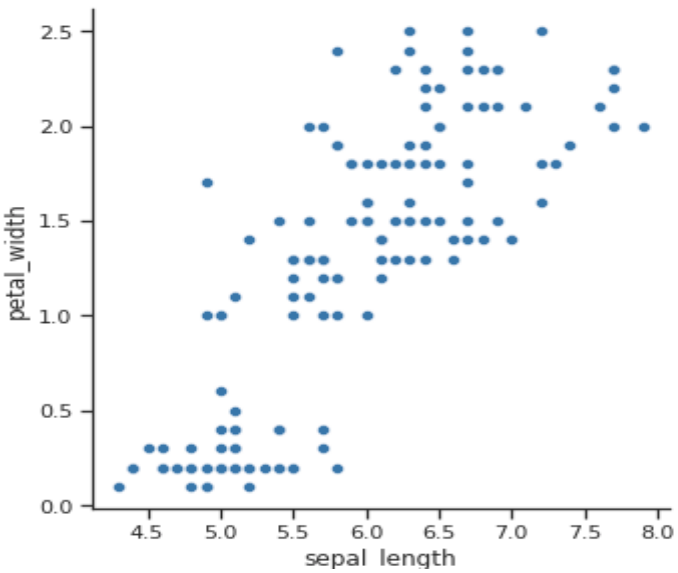
```
import seaborn as sns
sns.set(style ="ticks")
dataset = sns.load_dataset('iris')
col=['sepal_length','sepal_
sns.relplot(x ="sepal_length", y ="petal_width", data = dataset)
```

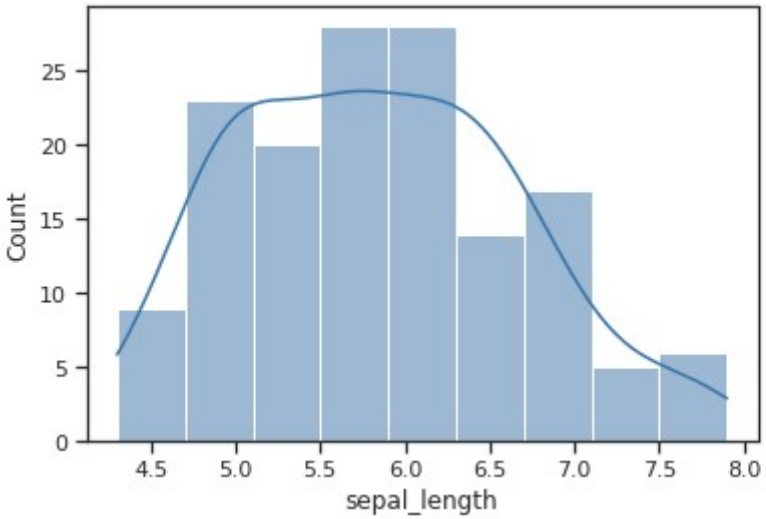
```
import numpy as np
import pandas as pd
import seaborn as sns
# Load dataset
dataset = sns.load_dataset("iris")
```

```
# Plot histogram
sns.histplot(data = dataset, x = "sepal_length", kde = True)
```

OUTPUT







Reference:

https://medium.com/@avulurivenkatasaireddy/exploratory-data-analysis-of-iris-data-set-using-python-823e54110d2d

https://web.ics.purdue.edu/~yrosokha/code/Seaborn_Example_1.html

https://www.section.io/engineering-education/seaborn-tutorial/

https://seaborn.pydata.org/introduction.html

https://towardsdatascience.com/seaborn-python-8563c3d0ad41