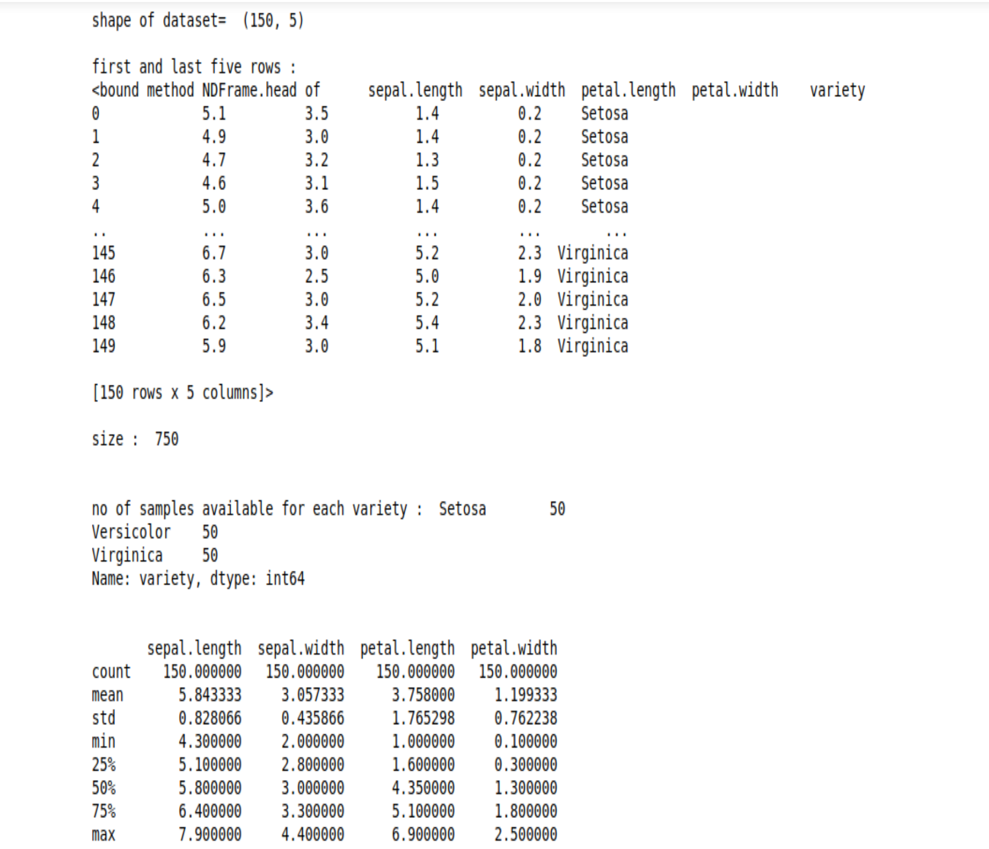
DATA HANDLING USING ‘Pandas’ and DATA VISUALIZATION USING ‘Seaborn’

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
iris=pd.read_csv("iris.csv")
print("shape of dataset= ",iris.shape)
print("")
print("first and last five rows : ")
print(iris.head)
print(" ")
print("size : ",iris.size)
print(" ")
print(" ")
print("no of samples available for each variety : ",
iris["variety"].value_counts())
print(" ")
print(" ")
print(iris.describe())
```

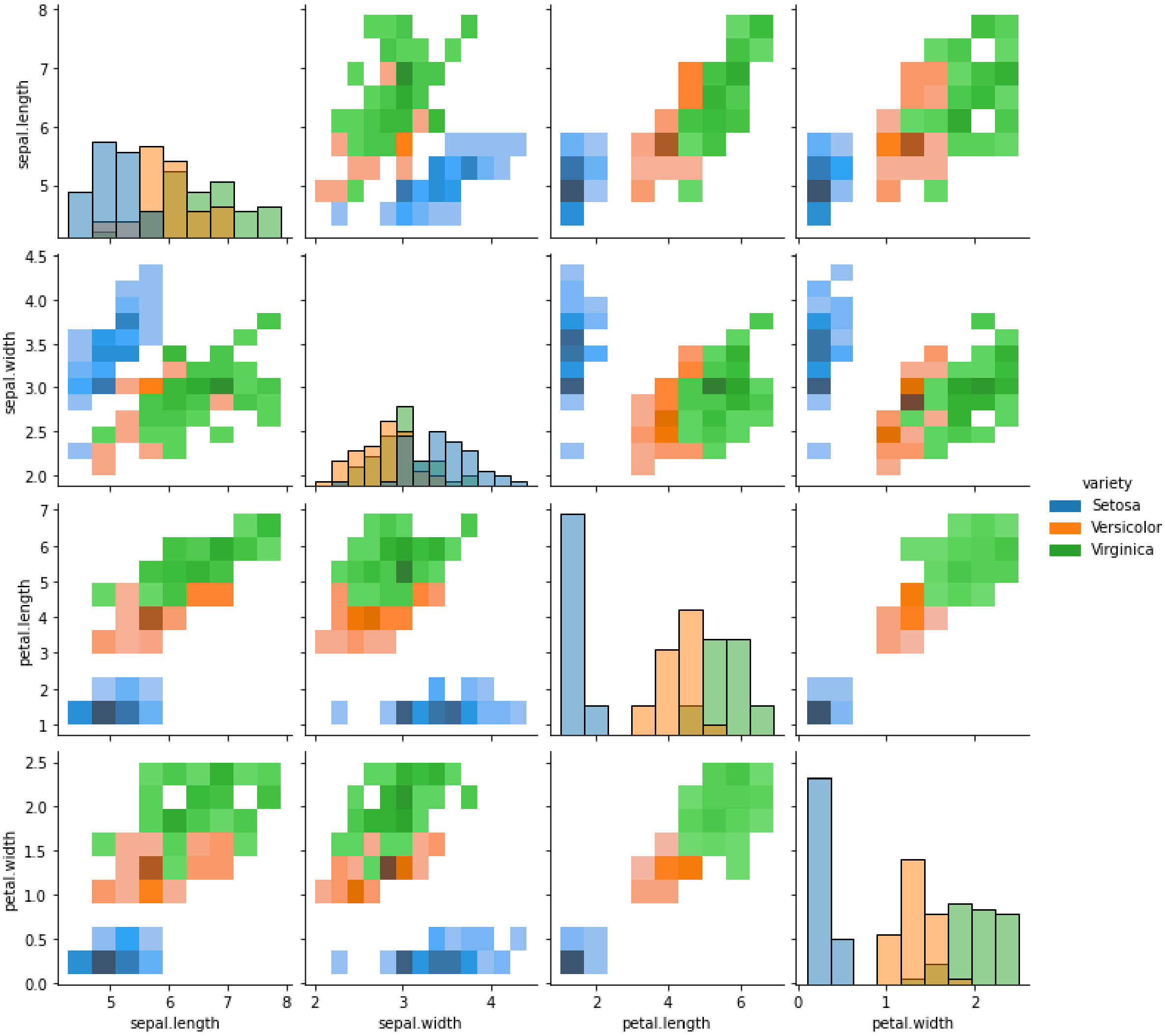


```
shape of dataset=  (150, 5)

first and last five rows :
<bound method NDFrame.head of      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
..            ...          ...           ...          ...        ...
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

[150 rows x 5 columns]>

size :  750


no of samples available for each variety :  Setosa        50
Versicolor    50
Virginica     50
Name: variety, dtype: int64


       sepal.length  sepal.width  petal.length  petal.width
count    150.000000   150.000000    150.000000   150.000000
mean       5.843333     3.057333      3.758000     1.199333
std        0.828066     0.435866      1.765298     0.762238
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
```

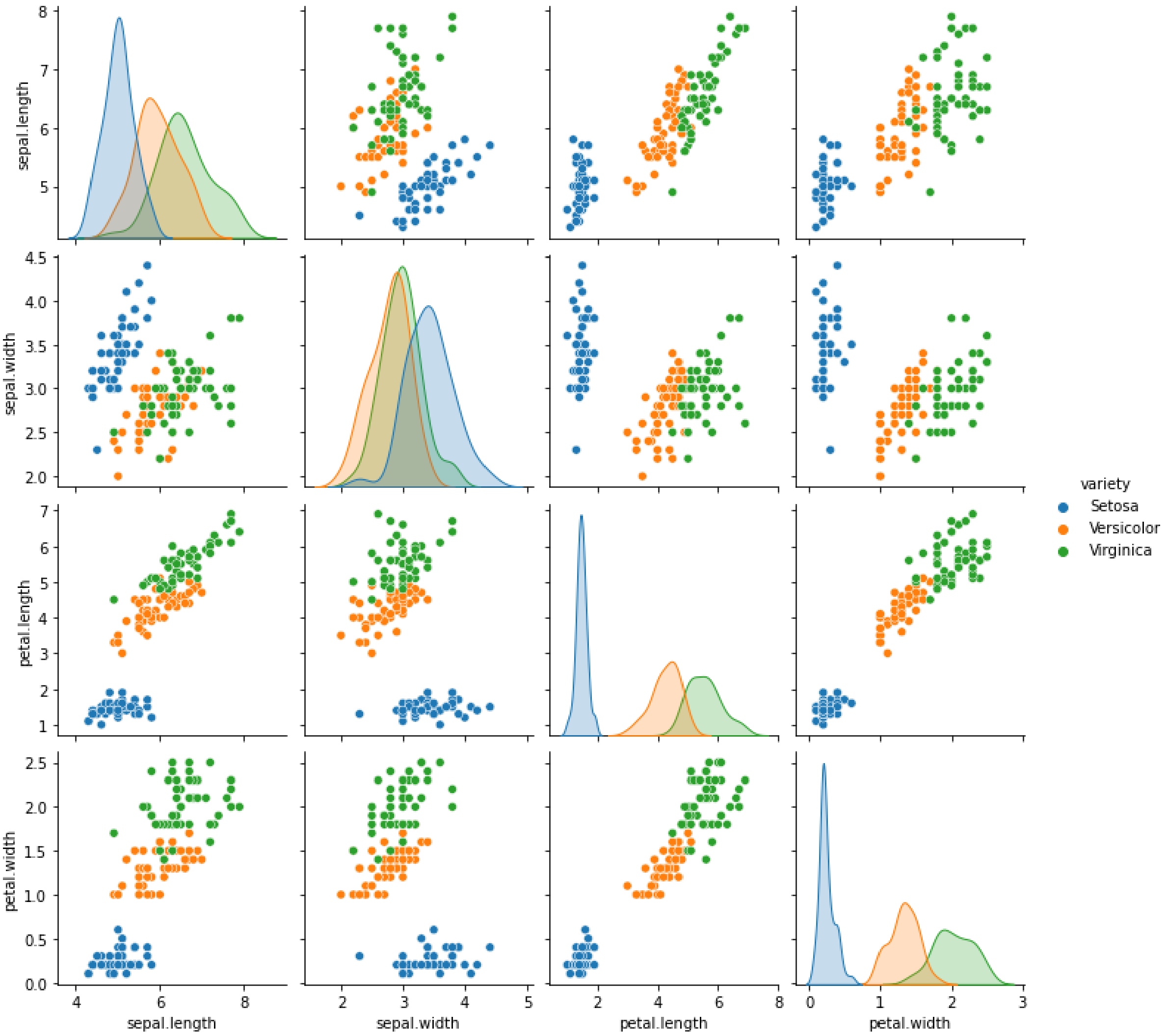
2. Use pairplot() function to display pairwise relationships between attributes. Try different kind of plots {***‘scatter’, ‘kde’, ‘hist’, ‘reg’***} and different kind of markers

code:

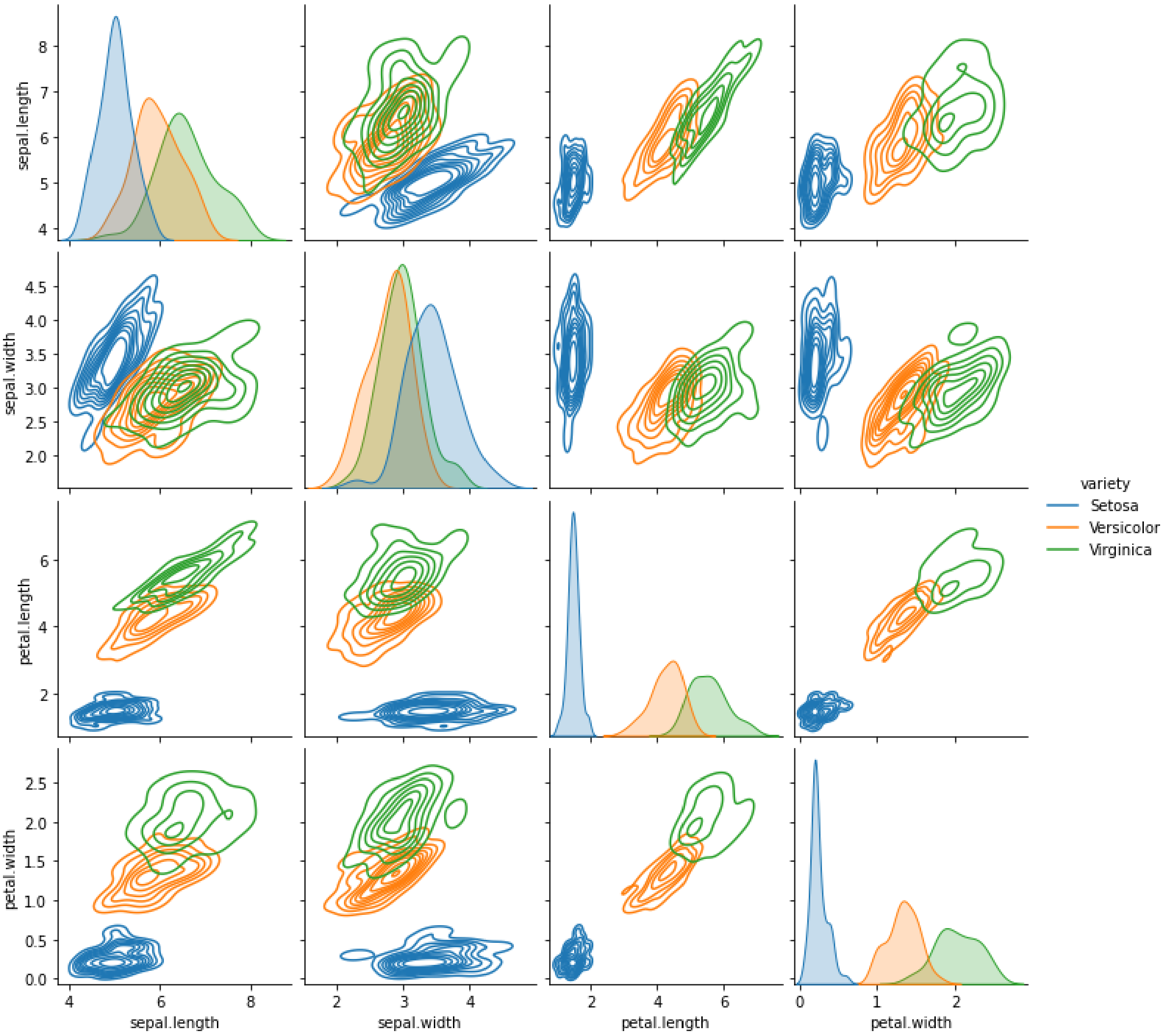
```
sns.pairplot(iris, hue="variety", kind="hist")
```



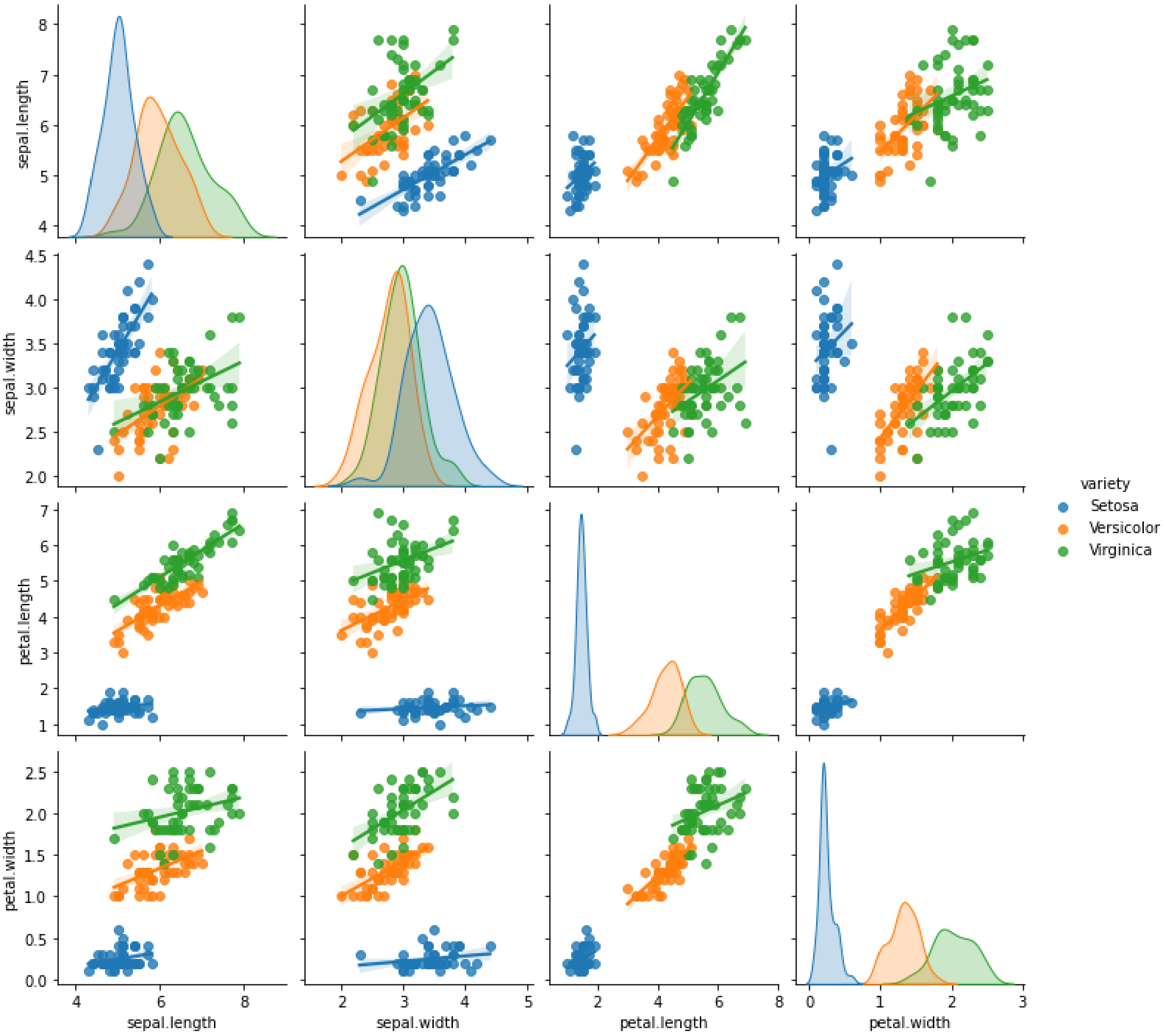
```
sns.pairplot(iris, hue="variety", kind="scatter")
```



```
sns.pairplot(iris, hue="variety", kind="kde")
```



```
sns.pairplot(iris, hue="variety", kind="reg")
```



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

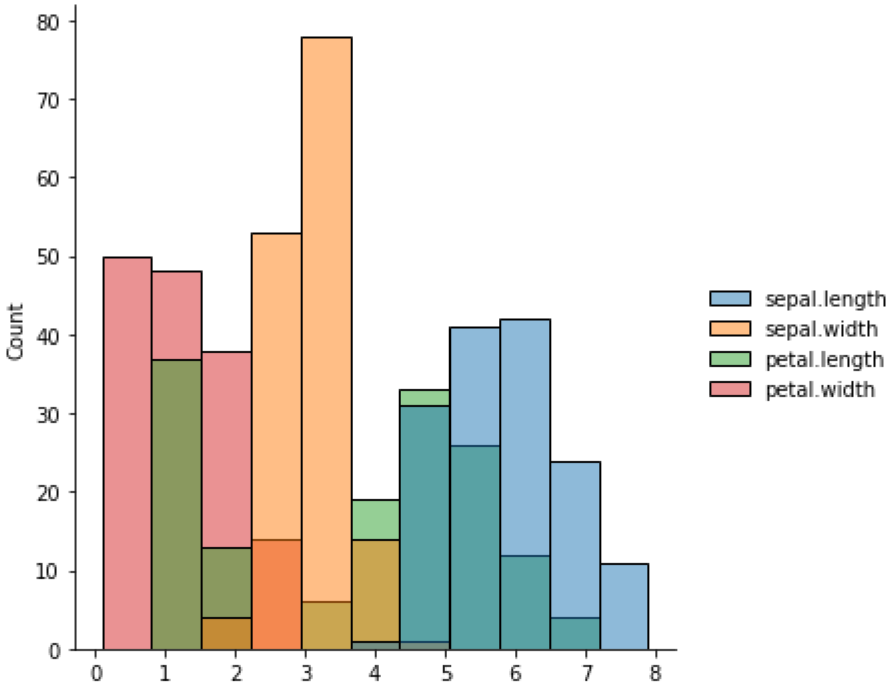
1)displot()

2) histplot()

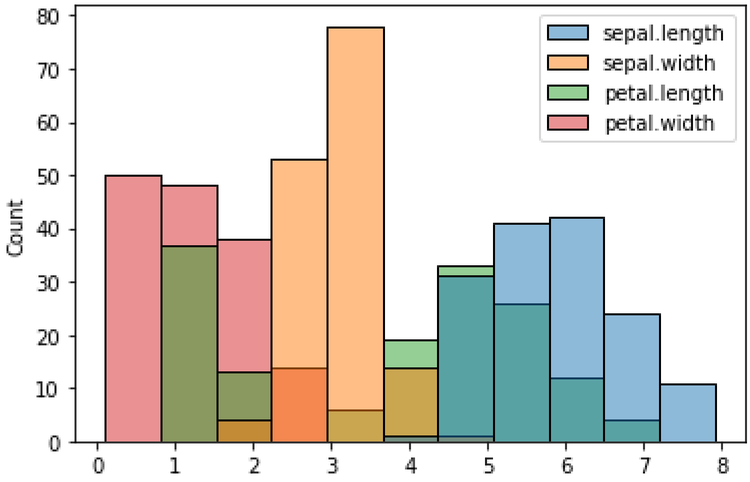
3) relplot()

code:

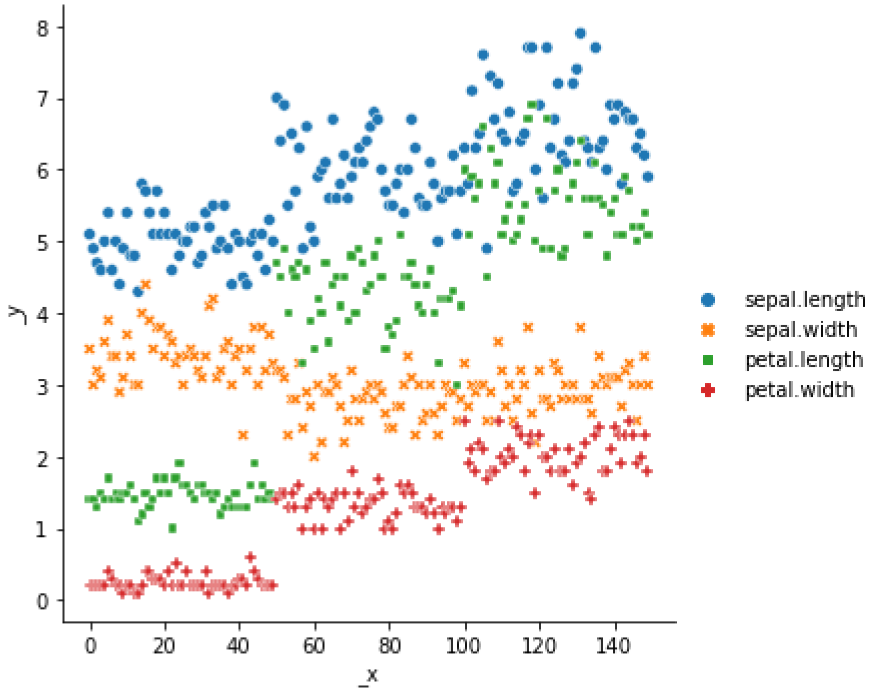
sns.displot(data=iris)



sns.histplot(data=iris)



sns.relplot(data=iris)


y
_x
sepal.length
sepal.width
petal.length
petal.width