Lab 8 13-09-2022

September 13, 2022

0.1 Programs using matplotlib / seaborn for data visualisation

0.1.1 a)Write a program to draw univariate visualization plots(line plot,histogram,boxplot,barchart,piechart) with matplotlib for iris dataset

[54]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

[9]:
```
i = pd.read_csv('iris.csv')
df = pd.DataFrame(i)
print(df)
```

```
     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
..            …            …             …            …        …
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]
```

.

.

.

.

.

.

.

.

.

.

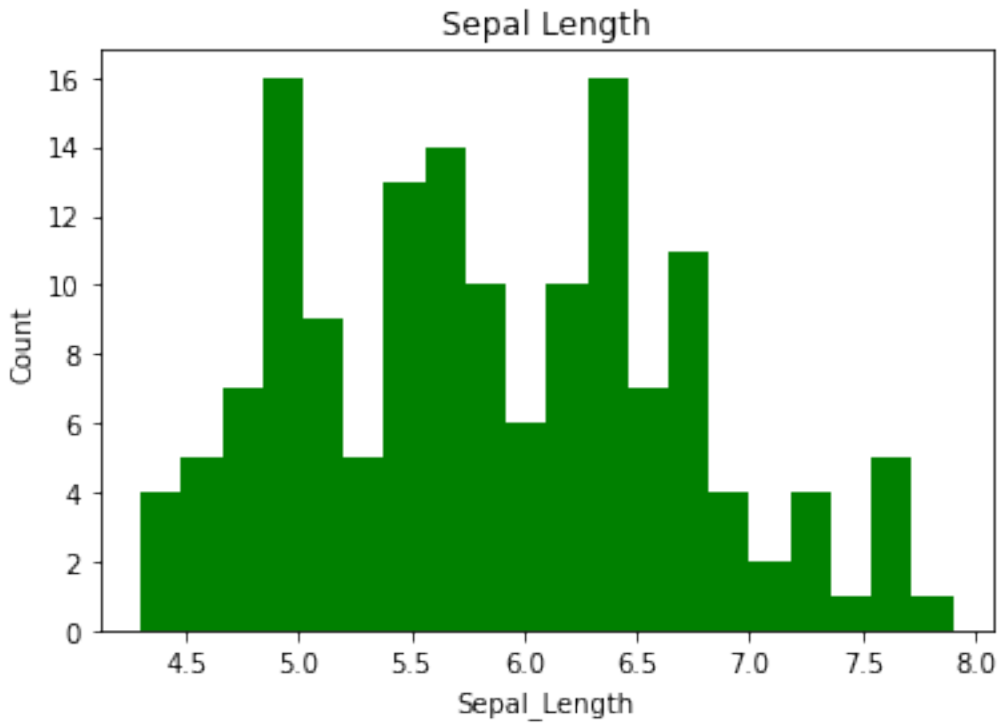
.

.

0.1.2 Histogram

[76]:
```
x = df["sepal_length"]
plt.hist(x, bins = 20, color = "g")
plt.title("Sepal Length")
plt.xlabel("Sepal_Length")
plt.ylabel("Count")
```

[76]:
```
Text(0, 0.5, 'Count')
```



.

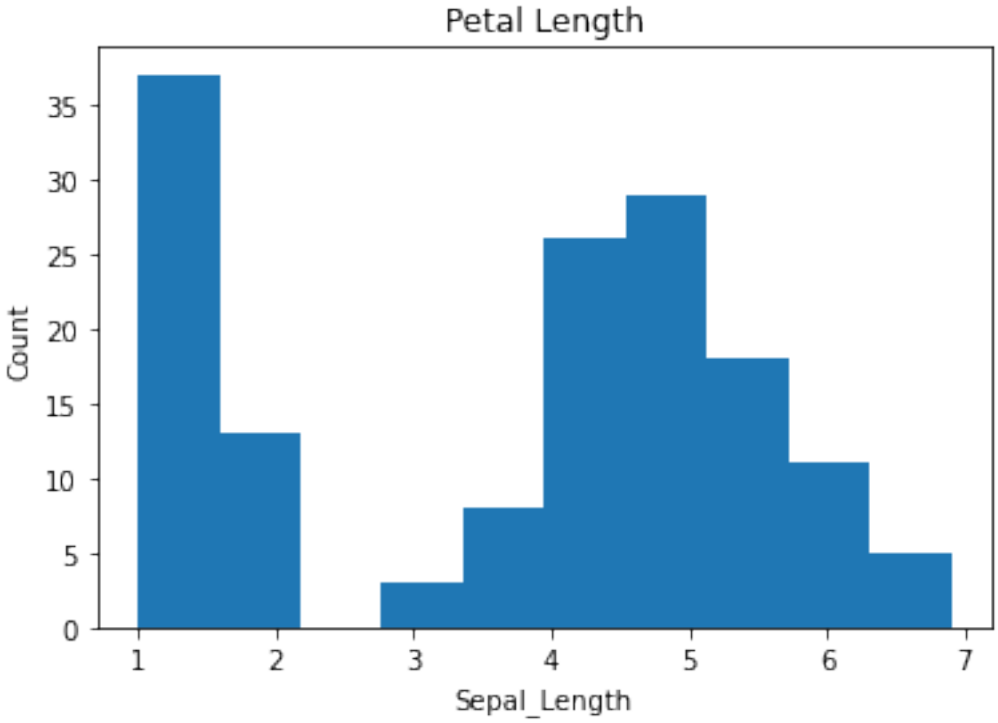
.

.

.

[69]:
```
df["petal_length"].plot(kind = 'hist')
plt.title("Petal Length")
plt.xlabel("Sepal_Length")
plt.ylabel("Count")
```

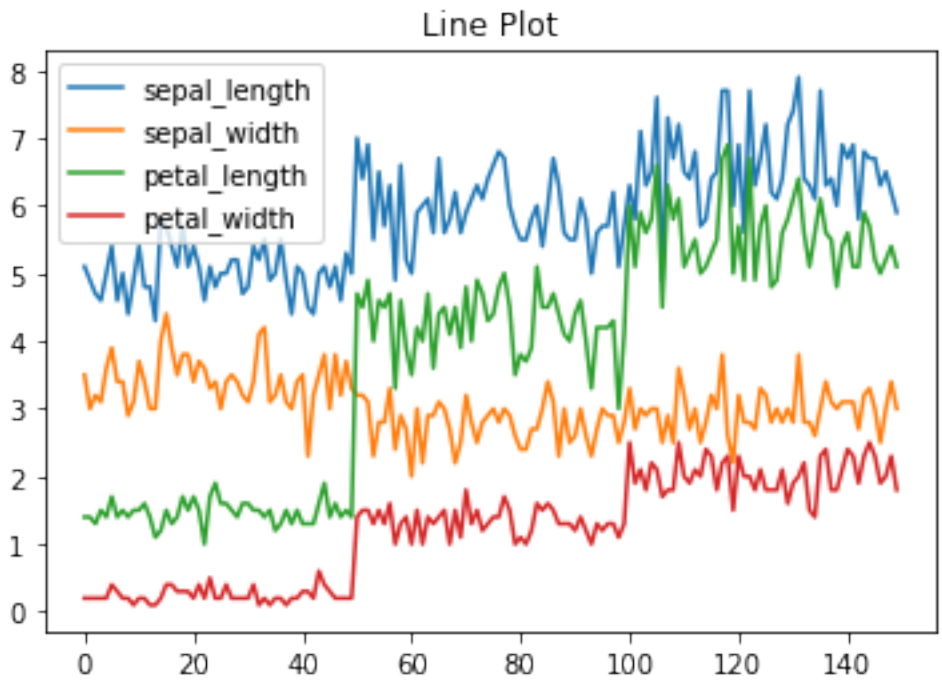
[69]:
```
Text(0, 0.5, 'Count')
```



0.1.3 Line

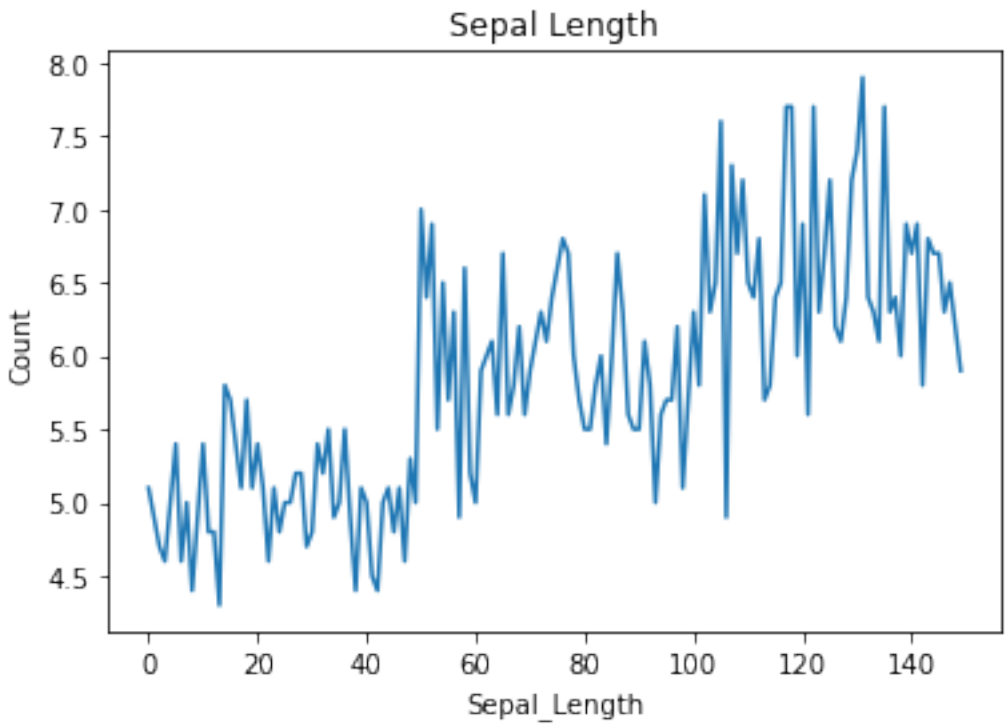
[70]:
```
df.plot(kind = 'line')
plt.title("Line Plot")
```

[70]:
```
Text(0.5, 1.0, 'Line Plot')
```



[21]:
```
df["sepal_length"].plot(kind = 'line')
plt.title("Sepal Length")
plt.xlabel("Sepal_Length")
plt.ylabel("Count")
```

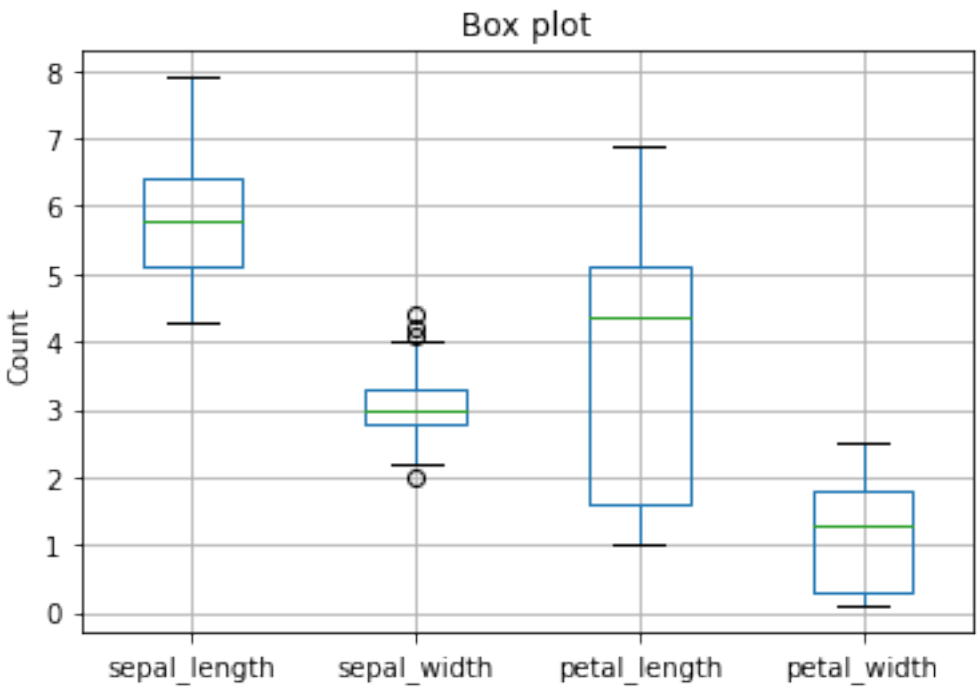
[21]:
```
Text(0, 0.5, 'Count')
```



0.1.4 Box

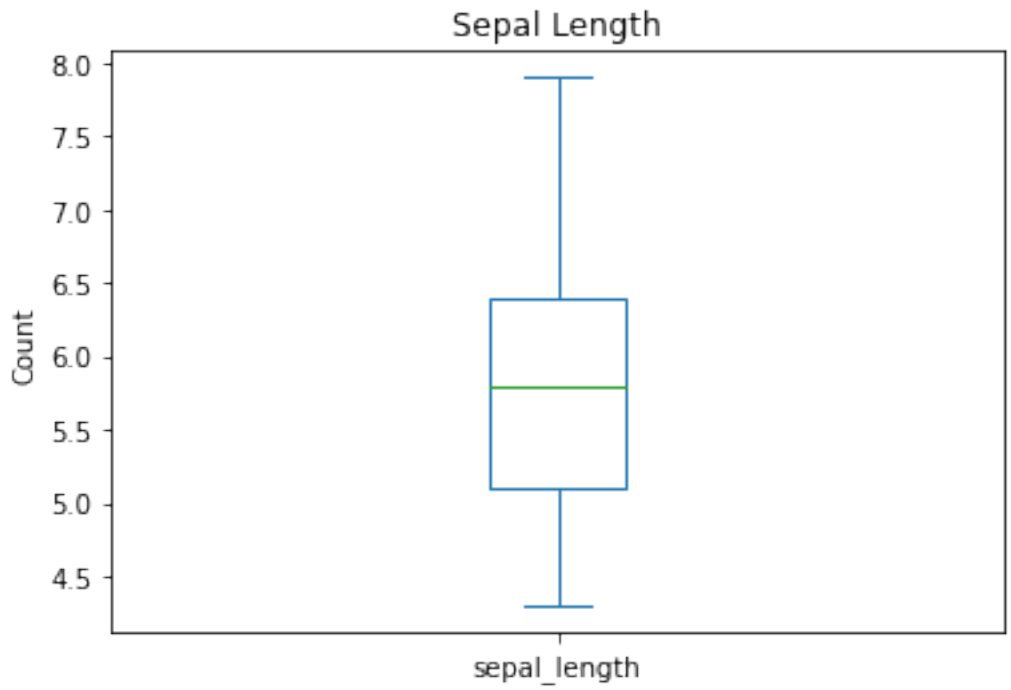
[73]:
```
df.boxplot()
plt.ylabel("Count")
plt.title("Box plot")
```

[73]:
```
Text(0.5, 1.0, 'Box plot')
```



[71]:
```
df["sepal_length"].plot(kind = 'box')
plt.title("Sepal Length")
#plt.xlabel("Sepal_Length")
plt.ylabel("Count")
```

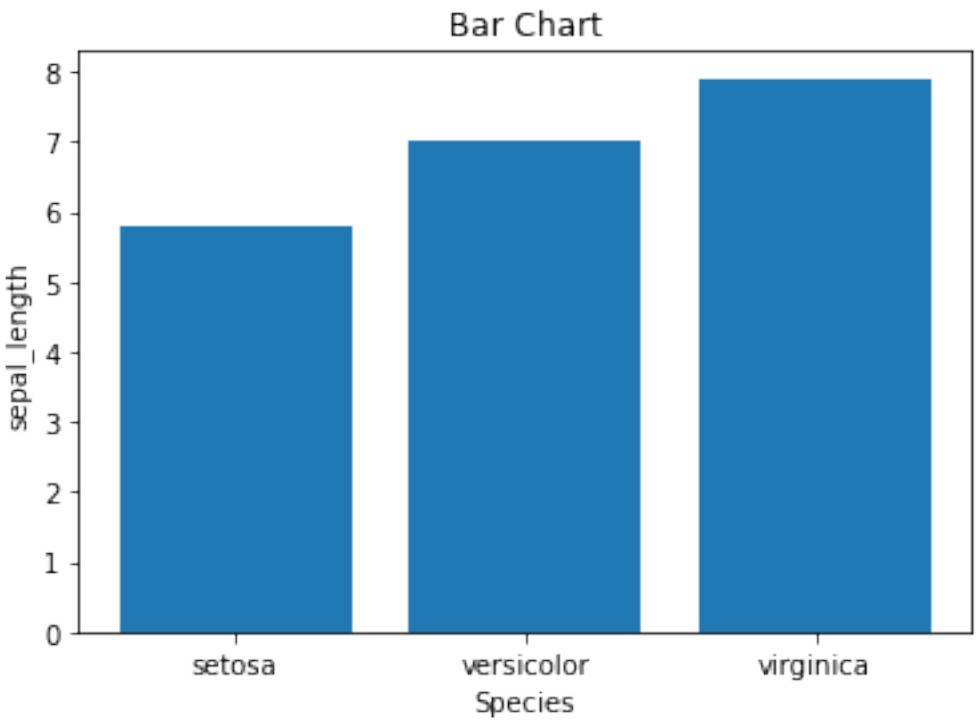
[71]:
```
Text(0, 0.5, 'Count')
```



0.1.5 Barplot

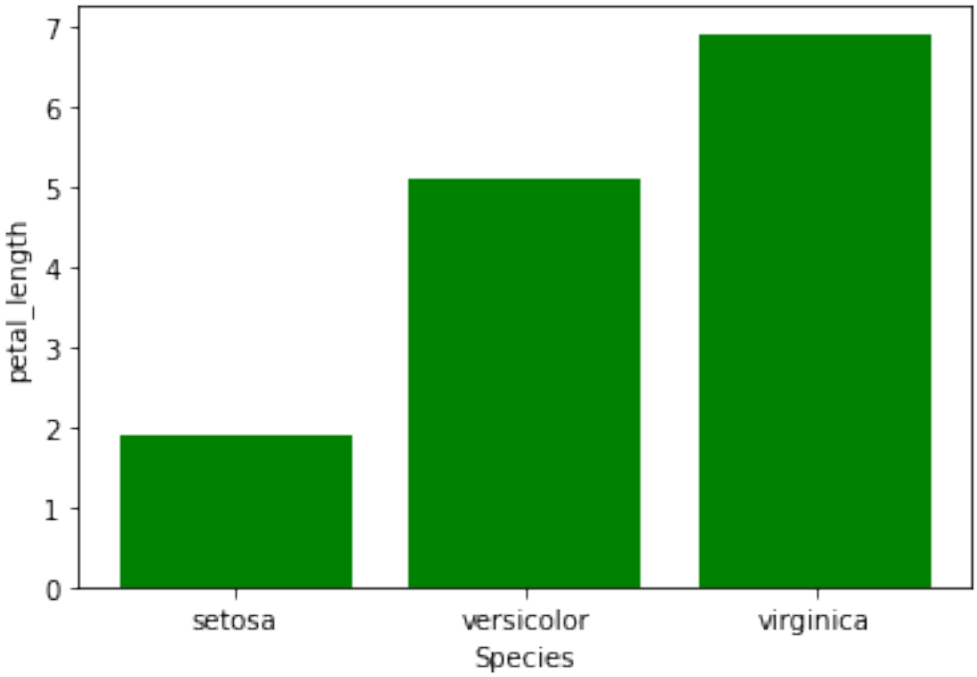
[77]:
```
plt.bar(df['species'],df['sepal_length'])
plt.title('Bar Chart')
plt.xlabel('Species')
plt.ylabel('sepal_length')
```

[77]:
```
Text(0, 0.5, 'sepal_length')
```



[96]:
```
plt.bar(df['species'],df['petal_length'], color = 'g')
plt.xlabel('Species')
plt.ylabel('petal_length')
```

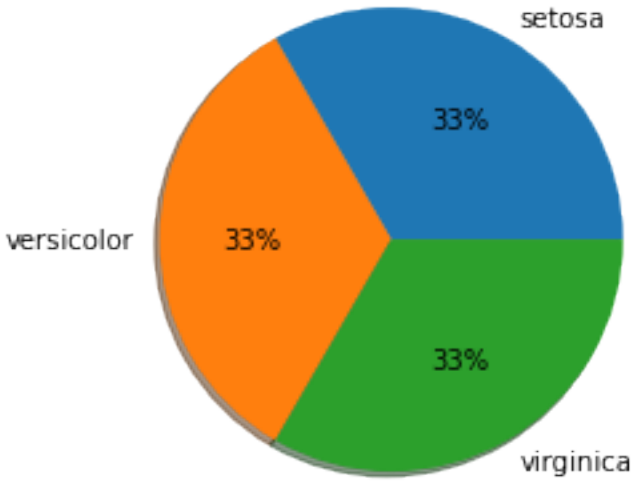
[96]:
```
Text(0, 0.5, 'petal_length')
```



0.1.6 Pie

[88]:
```
plt.pie(df['species'].
 ↪value_counts(),labels=['setosa','versicolor','virginica'],autopct='%1.0f%%',␣
 ↪shadow=True)
```

[88]:
```
([<matplotlib.patches.Wedge at 0x7fe312ad0e20>,
  <matplotlib.patches.Wedge at 0x7fe312adc7f0>,
  <matplotlib.patches.Wedge at 0x7fe312ae81c0>],
 [Text(0.5499999702695115, 0.9526279613277875, 'setosa'),
  Text(-1.0999999999999954, -1.0298943258065002e-07, 'versicolor'),
  Text(0.5500001486524352, -0.9526278583383436, 'virginica')],
 [Text(0.2999999837833699, 0.5196152516333385, '33%'),
  Text(-0.5999999999999974, -5.6176054134900006e-08, '33%'),
  Text(0.30000008108314646, -0.5196151954572783, '33%')])
```

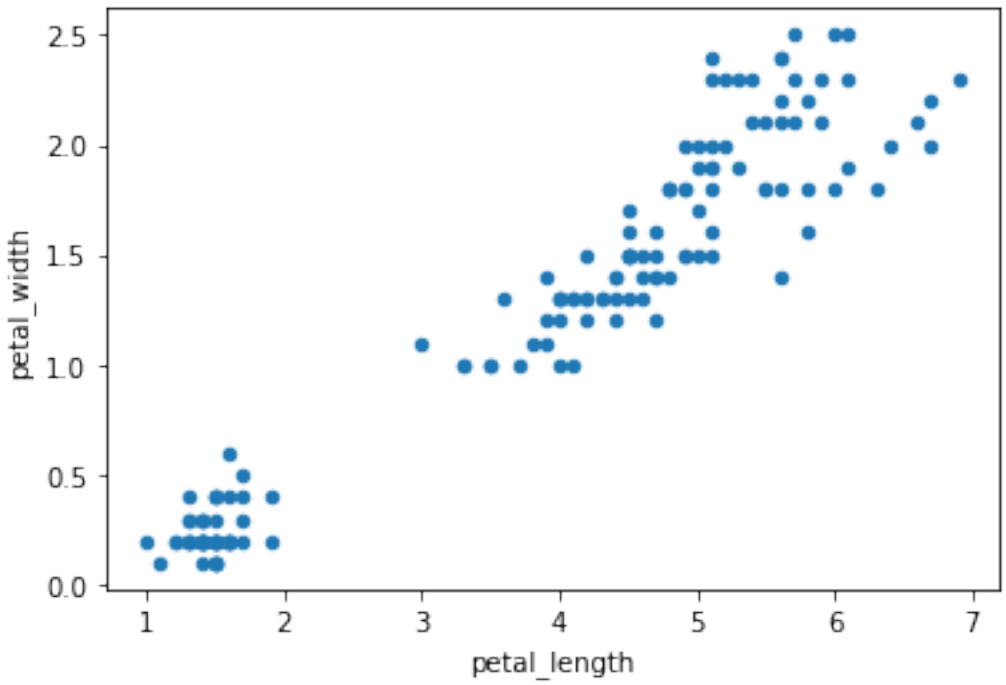


0.1.7 b)Write a program to draw multivariate visualization plots(scatter plots, scatter multiple,scatter matrix,bubble plot) with matplotlib for iris dataset

0.1.8 Scatter plot

[42]:
```
df.plot.scatter(x='petal_length', y = 'petal_width')
```

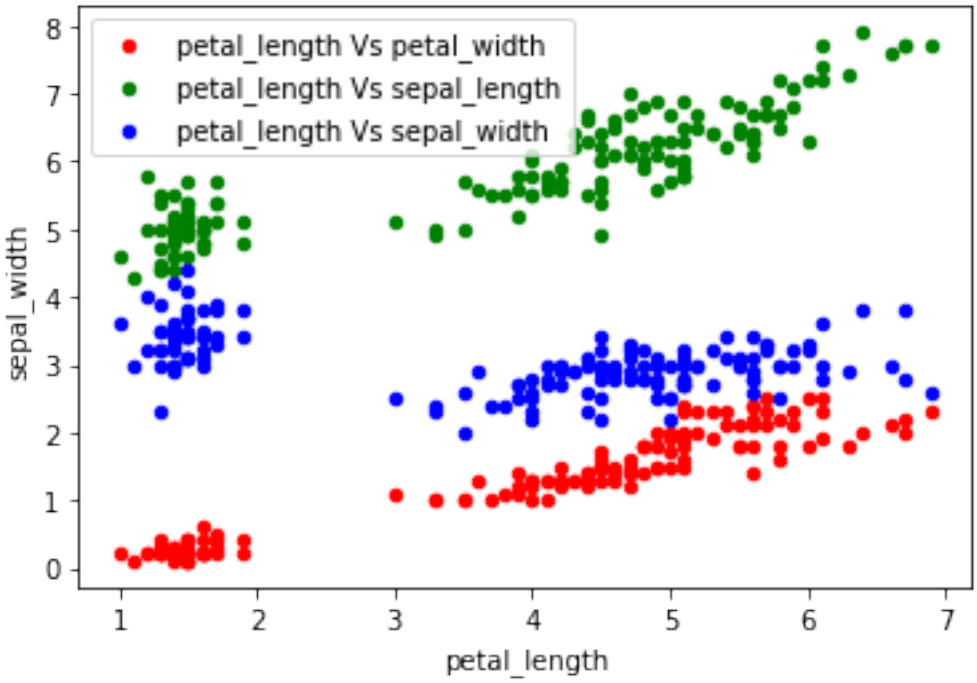
[42]:
```
<AxesSubplot:xlabel='petal_length', ylabel='petal_width'>
```



0.1.9 Scatter Multiple

[44]:
```
ax1 = df.plot(kind='scatter', x='petal_length', y = 'petal_width',␣
 ↪label='petal_length Vs petal_width', color='r')
ax2 = df.plot(kind='scatter', x='petal_length', y = 'sepal_length',␣
 ↪label='petal_length Vs sepal_length',color='g',ax=ax1)
ax3 = df.plot(kind='scatter',  x='petal_length', y =␣
 ↪'sepal_width',label='petal_length Vs sepal_width', color='b',ax=ax2)
print(ax1 == ax2 == ax3)
```

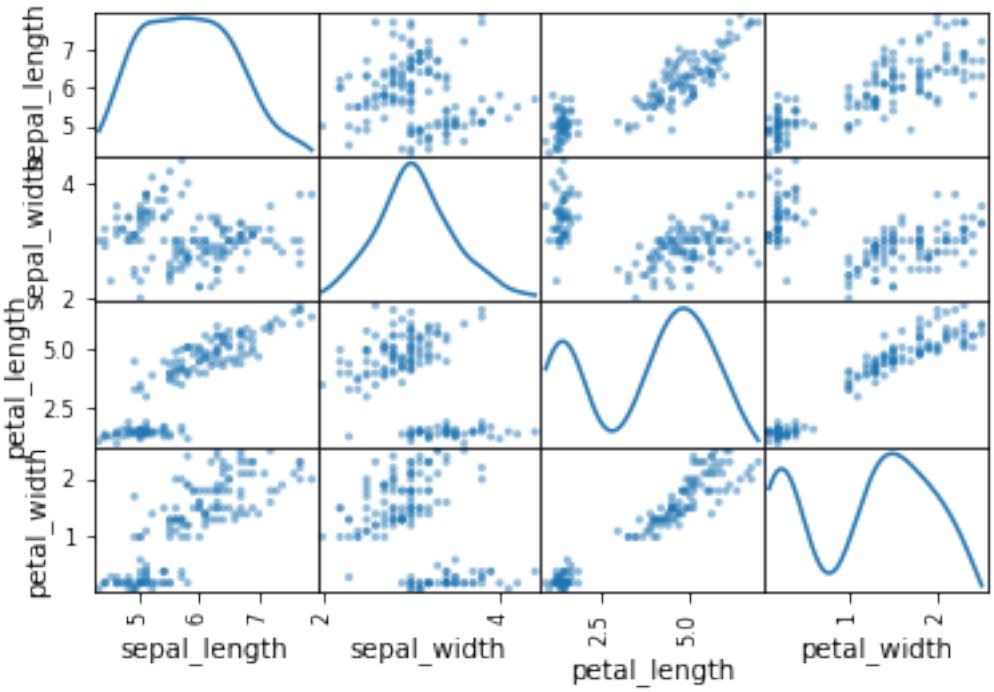
```
True
```

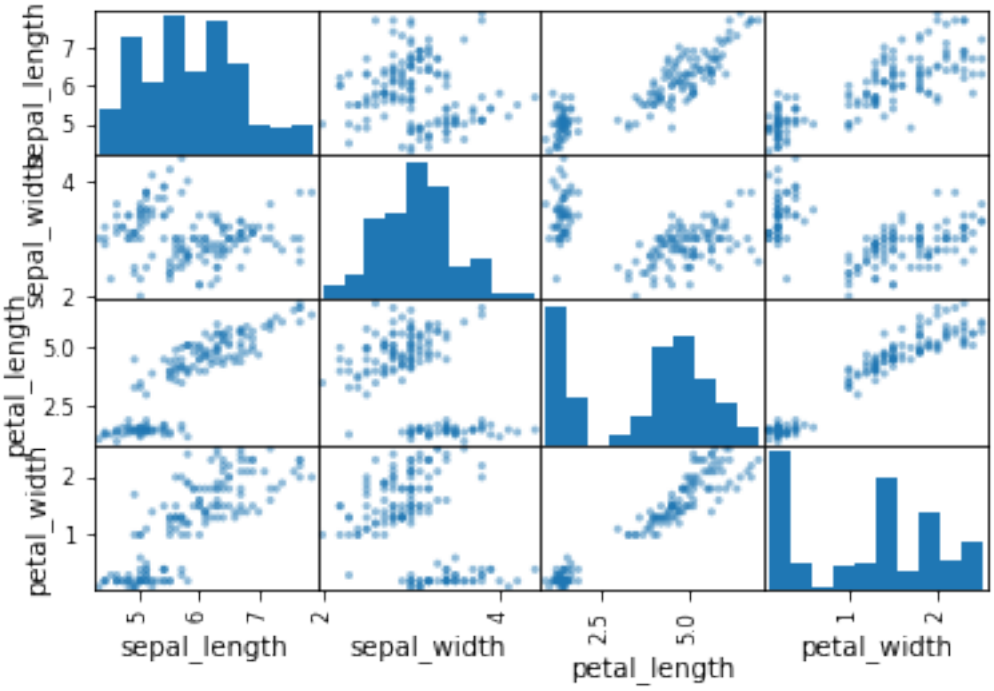


0.1.10 Scatter Matrix

[47]:
```
from pandas.plotting import scatter_matrix
scatter_matrix(df, alpha=0.5, diagonal='kde')
pd.plotting.scatter_matrix(df)
```

[47]:
```
array([[<AxesSubplot:xlabel='sepal_length', ylabel='sepal_length'>,
        <AxesSubplot:xlabel='sepal_width', ylabel='sepal_length'>,
        <AxesSubplot:xlabel='petal_length', ylabel='sepal_length'>,
        <AxesSubplot:xlabel='petal_width', ylabel='sepal_length'>],
       [<AxesSubplot:xlabel='sepal_length', ylabel='sepal_width'>,
        <AxesSubplot:xlabel='sepal_width', ylabel='sepal_width'>,
        <AxesSubplot:xlabel='petal_length', ylabel='sepal_width'>,
        <AxesSubplot:xlabel='petal_width', ylabel='sepal_width'>],
       [<AxesSubplot:xlabel='sepal_length', ylabel='petal_length'>,
        <AxesSubplot:xlabel='sepal_width', ylabel='petal_length'>,
        <AxesSubplot:xlabel='petal_length', ylabel='petal_length'>,
        <AxesSubplot:xlabel='petal_width', ylabel='petal_length'>],
       [<AxesSubplot:xlabel='sepal_length', ylabel='petal_width'>,
        <AxesSubplot:xlabel='sepal_width', ylabel='petal_width'>,
        <AxesSubplot:xlabel='petal_length', ylabel='petal_width'>,
        <AxesSubplot:xlabel='petal_width', ylabel='petal_width'>]],
      dtype=object)
```

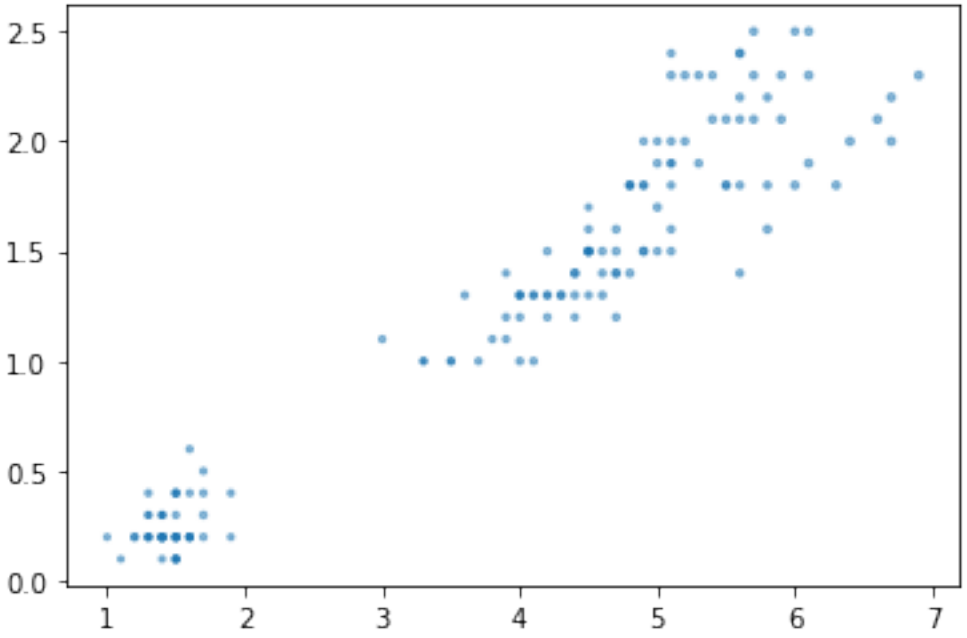

sepal_length
sepal_width
petal_length
petal_width
7
6
5
4
2
5.0
2.5
2
1
5
6
7
2
4
2.5
5.0
1
2


sepal_length
sepal_width
petal_length
petal_width
7
6
5
4
2
5.0
2.5
2
1
5
6
7
2
4
2.5
5.0
1
2

0.1.11 Bubble Plot

[98]:
```
plt.scatter('petal_length', 'petal_width',  s='sepal_length',  alpha=0.5, ␣
↪data=df)
```

[98]:
```
<matplotlib.collections.PathCollection at 0x7fe3123a3b50>
```

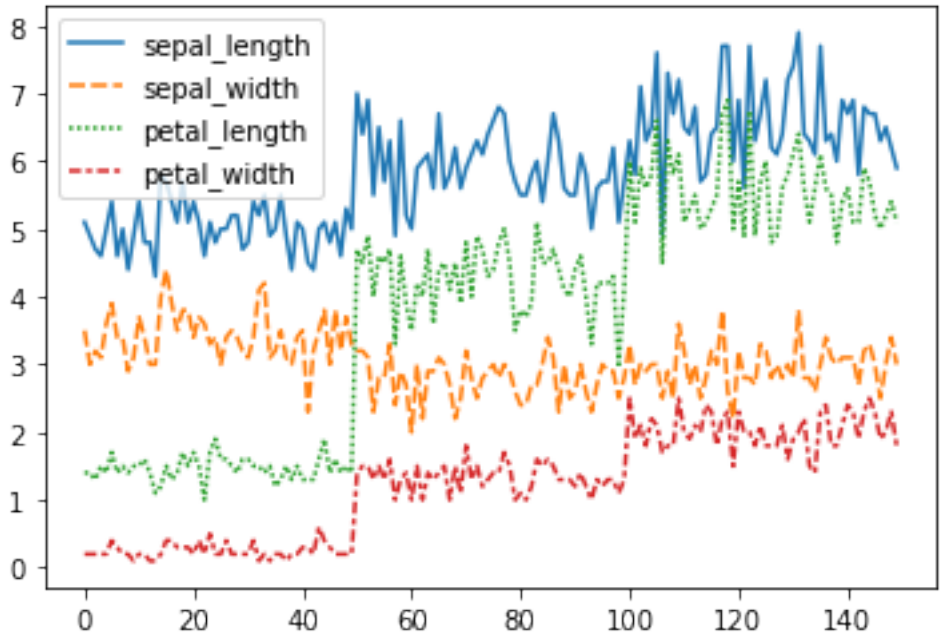


0.1.12 c)Write a program to draw univariate and multivariate visualization plots with seaborn(line plot, pairplot,jointplot,heatmap) for iris dataset

0.1.13 Line plot

[57]:
```
sns.lineplot(data=df)
```

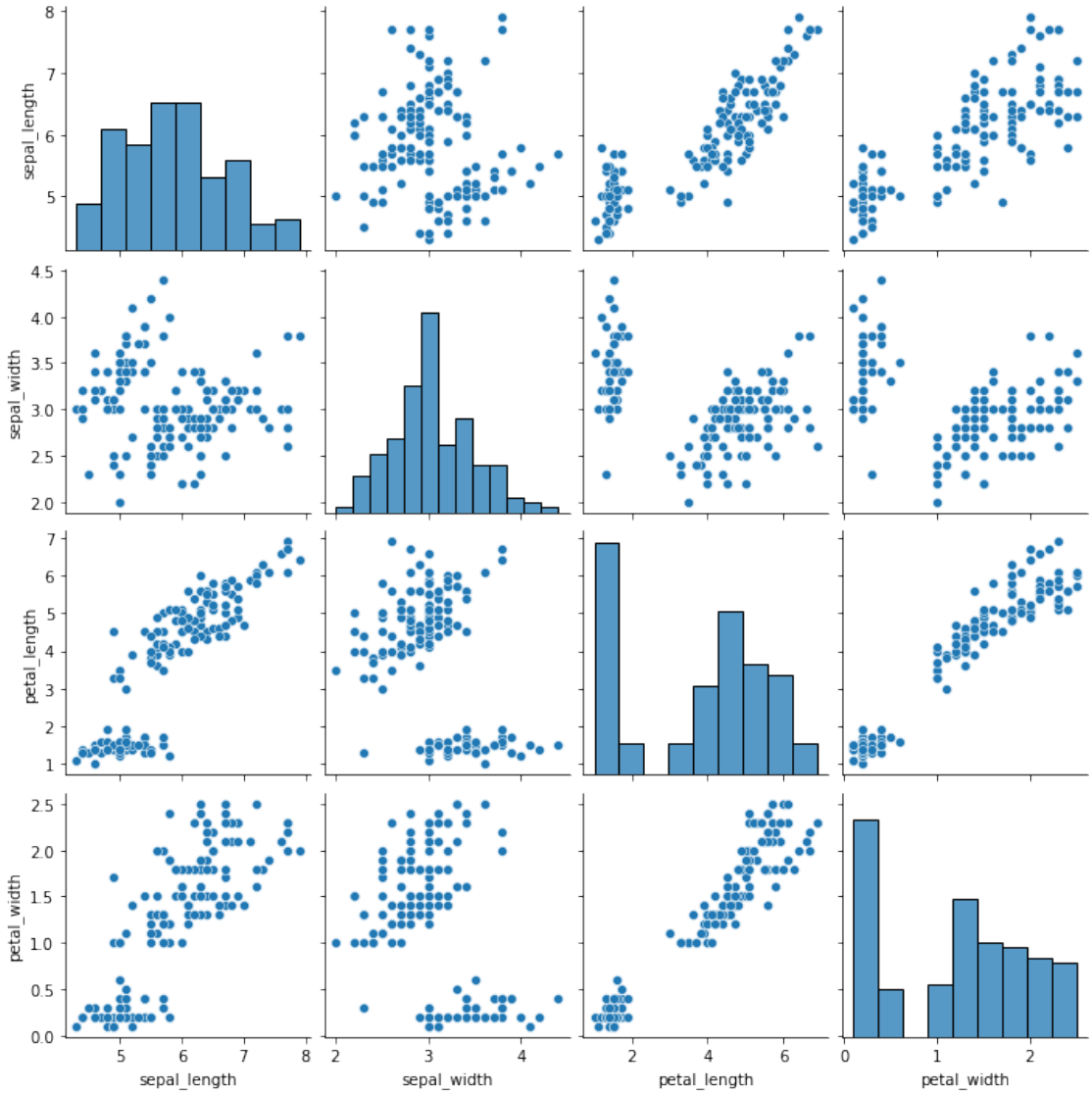
[57]:
```
<AxesSubplot:>
```



0.1.14 Pair Plot

[59]:
```
sns.pairplot(data=df,kind='scatter')
```

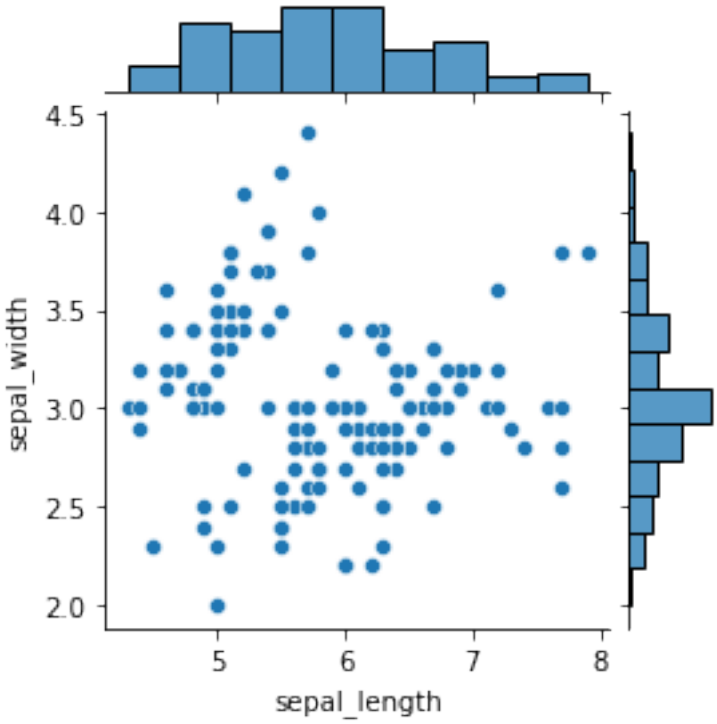
[59]:
```
<seaborn.axisgrid.PairGrid at 0x7fe314b0e490>
```



0.1.15 Joint Plot

[58]:
```
sns.jointplot(x='sepal_length',y='sepal_width',data=df,height=4)
```

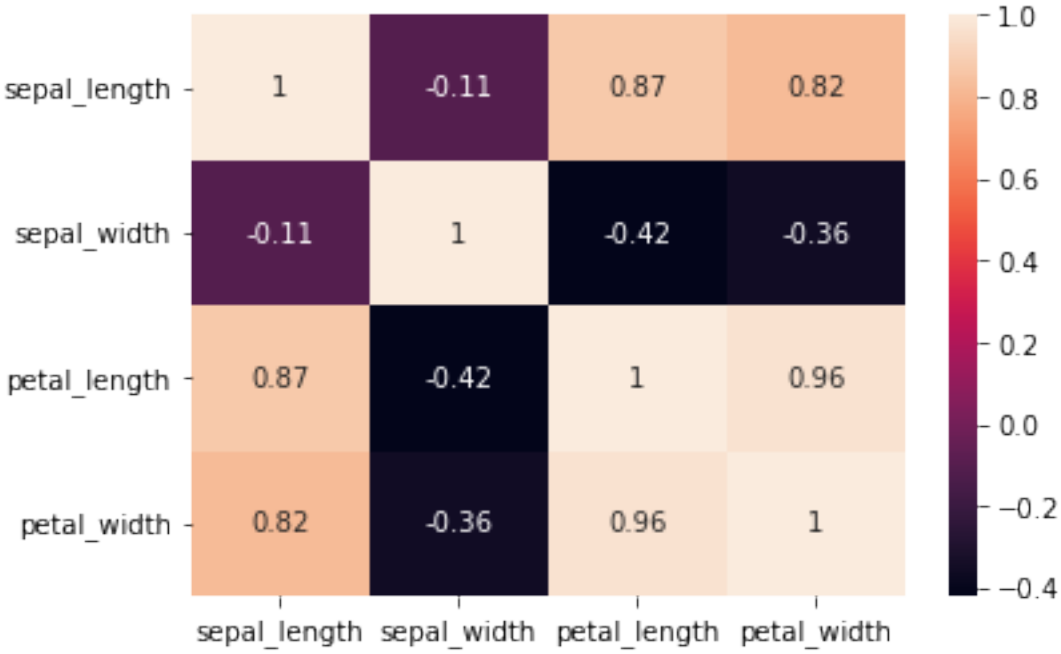
[58]:
```
<seaborn.axisgrid.JointGrid at 0x7fe314c604c0>
```



0.1.16 Heat Map

[63]:
```
sns.heatmap(df.corr(),annot=True)
```

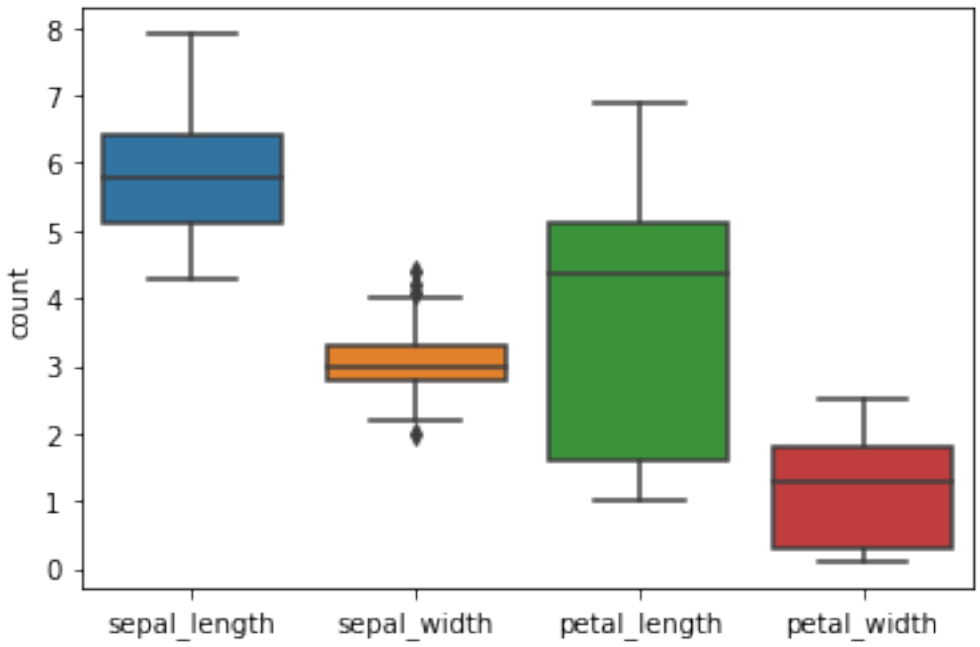
[63]: <AxesSubplot:>



0.1.17 Boxplot

[91]:
```
p = sns.boxplot(data=iris)
p.set_ylabel("count")
```

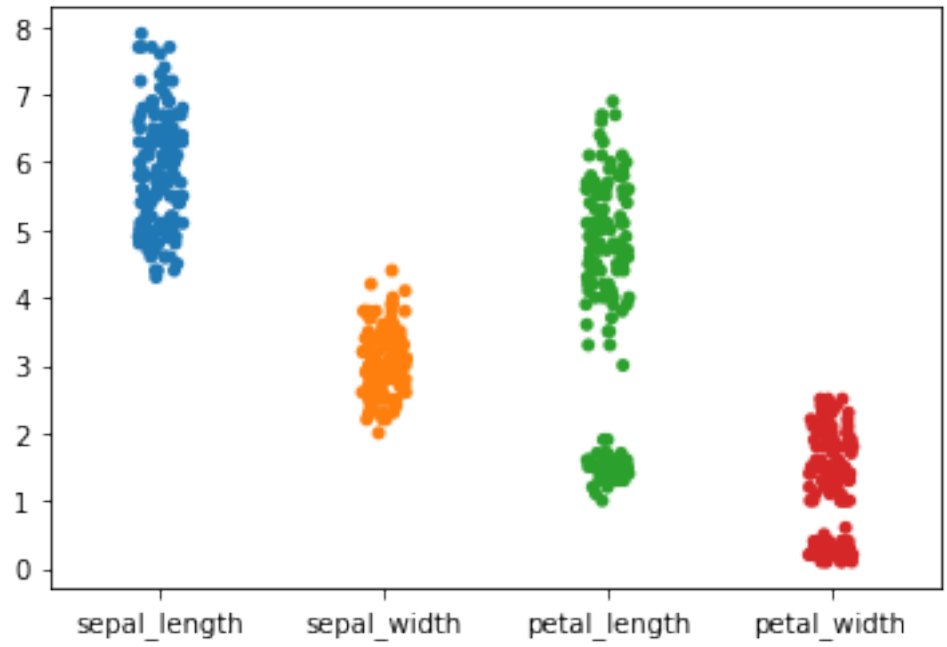
[91]:
```
Text(0, 0.5, 'count')
```



0.1.18 Striplot

[99]:
```
sns.stripplot(data=df)
```

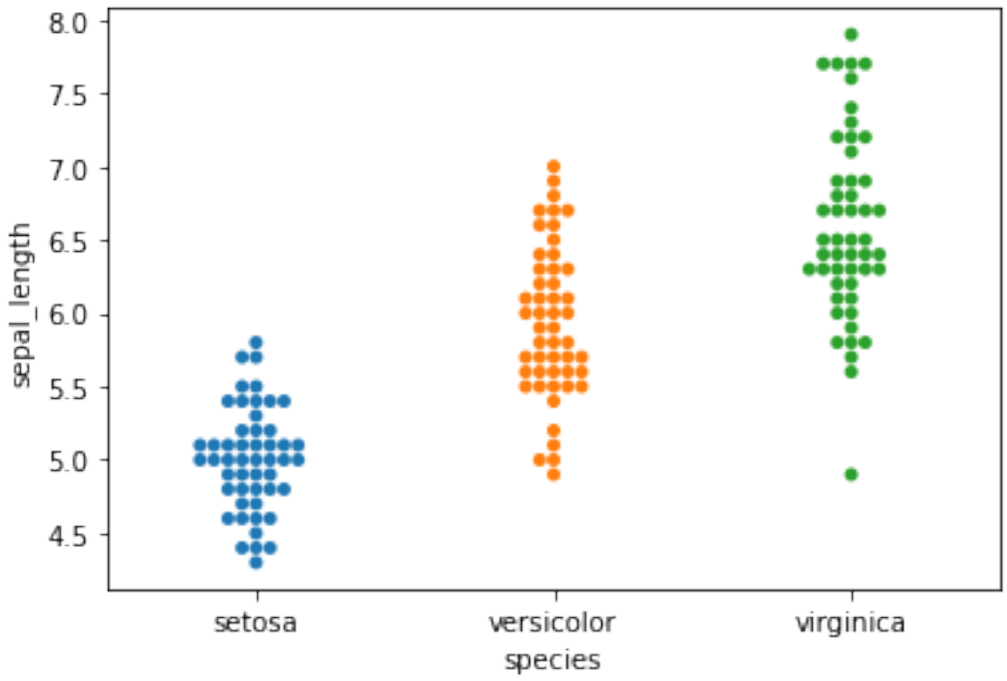
[99]:
```
<AxesSubplot:>
```



0.1.19 Swarm

[100]:
```
sns.swarmplot(x='species', y='sepal_length', data=df)
```

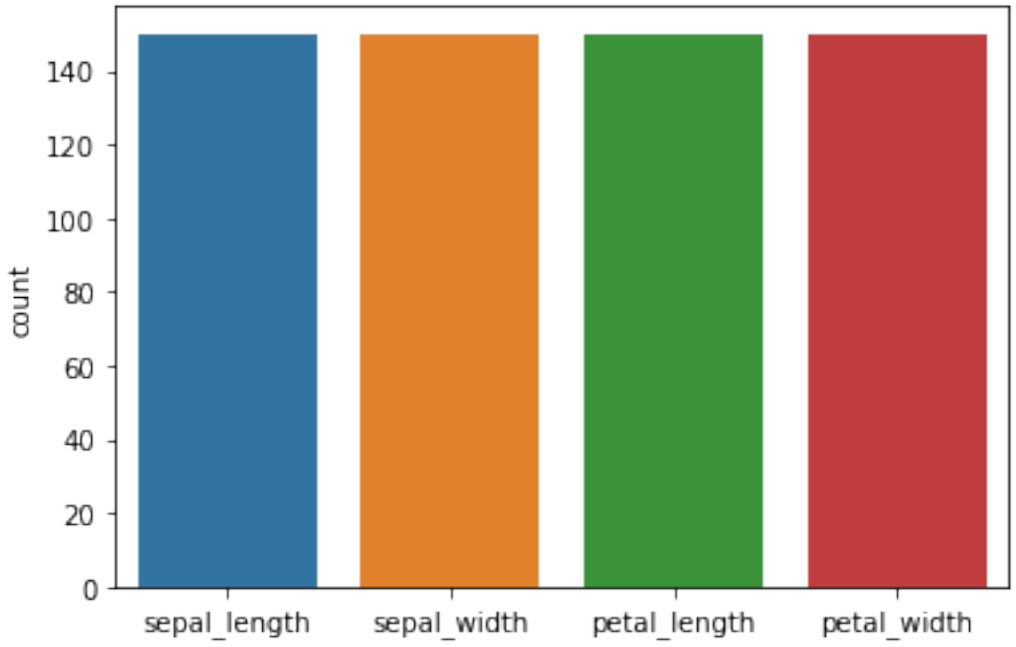
[100]:
```
<AxesSubplot:xlabel='species', ylabel='sepal_length'>
```



0.1.20 Count plot

[101]:
```
sns.countplot(data=df)
```

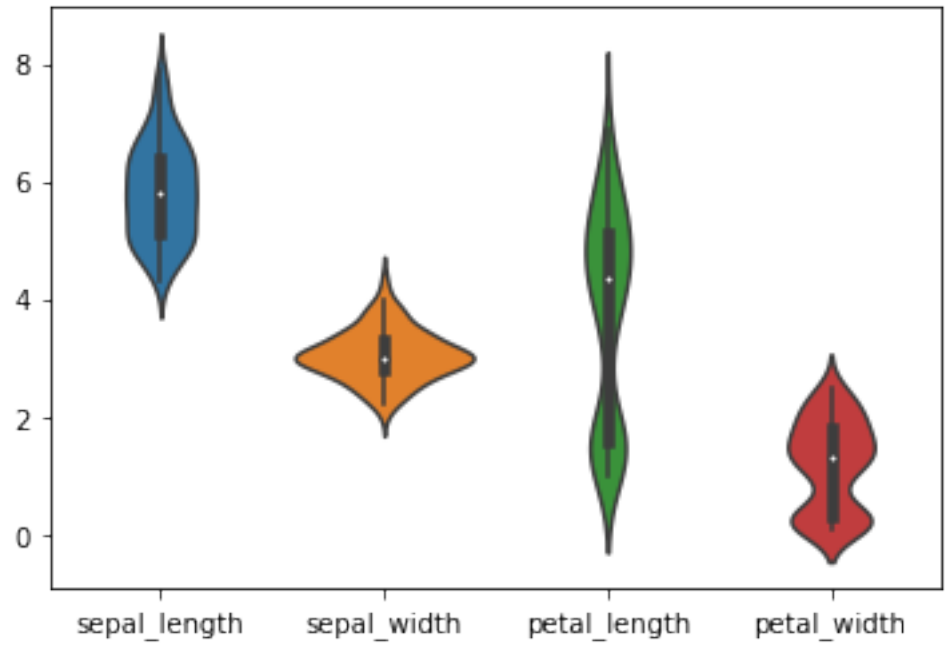
[101]: <AxesSubplot:ylabel='count'>



0.1.21 Violin plot

[102]:
```
sns.violinplot(data=df)
```

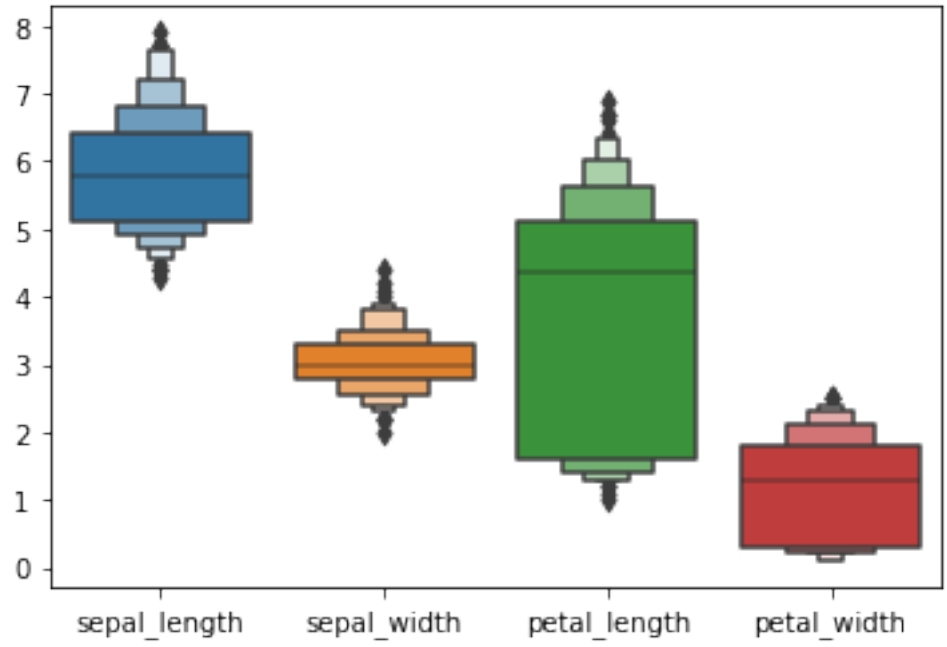
[102]: <AxesSubplot:>



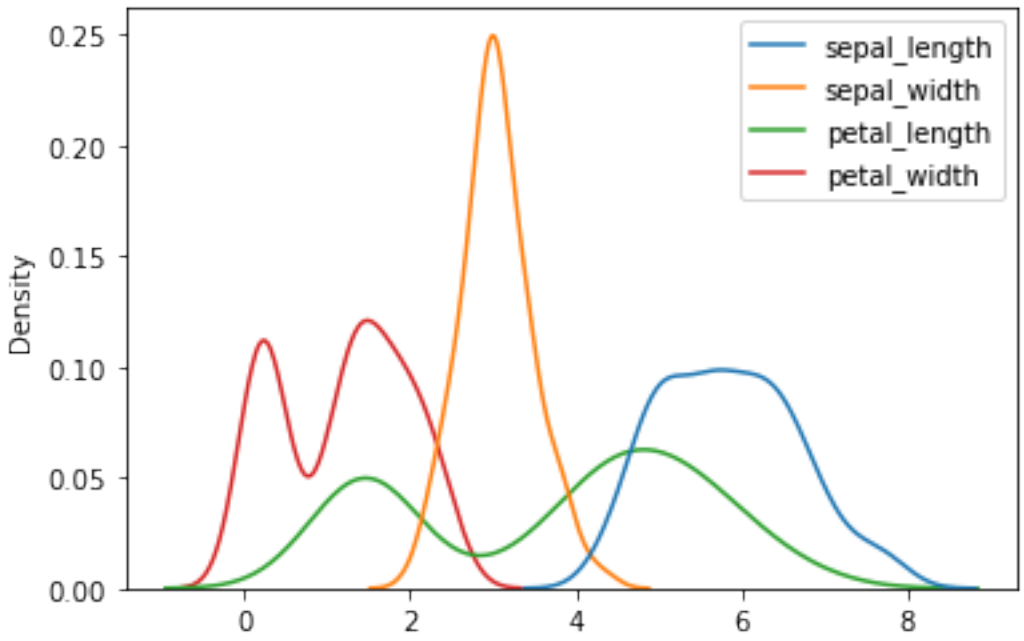
0.1.22 Boxenplot

[104]:
```
sns.boxenplot(data=df)
```

[104]: <AxesSubplot:>



0.1.23 kdeplot

[107]:
```
sns.kdeplot(data=df)
```

[107]: <AxesSubplot:ylabel='Density'>