Iris_analysis_results_graphs

August 13, 2025

Connected to Python 3.13.6

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

[ ]:
```
import numpy as np
```

[ ]:
```
import matplotlib.pyplot as plt
```

[ ]:
```
import seaborn as sns
```

[ ]:
```
import os
```

[ ]:
```
os.getcwd()
```

[ ]:
```
'c:\\Users\\HP\\Downloads\\datascience.py'
```

[ ]:
```
os.listdir('c:\\Users\\HP\\Downloads\\datascience.py')
```

[ ]:
```
['data.py',
 'data_files',
 'practice1.py',
 'practice2.py',
 'results_of_code.ipynb']
```

[ ]:
```
os.listdir('c:\\Users\\HP\\Downloads\\datascience.py\\data_files')
```

[ ]:
```
['IRIS.csv', 'irisdata.csv']
```

[ ]:
```
link_to_Iris_data = "c:\\Users\\HP\\Downloads\\datascience.
↪py\\data_files\\irisdata.csv"
```

[ ]:
```
link_to_Iris_data
```

[ ]:
```
'c:\\Users\\HP\\Downloads\\datascience.py\\data_files\\irisdata.csv'
```

[ ]:
```
Iris_data = pd.read_csv(link_to_Iris_data)
```

[ ]:
```
Iris_data.head(6)
```

[ ]:

	sepal_length	sepal_width	petal_length	petal_width	species
0	5.1	3.5	1.4	0.2	Iris-setosa
1	4.9	3.0	1.4	0.2	Iris-setosa
2	4.7	3.2	1.3	0.2	Iris-setosa
3	4.6	3.1	1.5	0.2	Iris-setosa
4	5.0	3.6	1.4	0.2	Iris-setosa
5	5.4	3.9	1.7	0.4	Iris-setosa

[ ]:
```
Iris_data.tail(6)
```

[ ]:

	sepal_length	sepal_width	petal_length	petal_width	species
144	6.7	3.3	5.7	2.5	Iris-virginica
145	6.7	3.0	5.2	2.3	Iris-virginica
146	6.3	2.5	5.0	1.9	Iris-virginica
147	6.5	3.0	5.2	2.0	Iris-virginica
148	6.2	3.4	5.4	2.3	Iris-virginica
149	5.9	3.0	5.1	1.8	Iris-virginica

[ ]:
```
Iris_data.shape
```

[ ]:
```
(150, 5)
```

[ ]:
```
Iris_data.dtypes
```

[ ]:
```
sepal_length    float64
sepal_width     float64
petal_length    float64
petal_width     float64
species          object
dtype: object
```

[ ]:
```
Iris_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 150 entries, 0 to 149
Data columns (total 5 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   sepal_length  150 non-null    float64
 1   sepal_width   150 non-null    float64
 2   petal_length  150 non-null    float64
 3   petal_width   150 non-null    float64
 4   species       150 non-null    object
dtypes: float64(4), object(1)
memory usage: 6.0+ KB
```

[ ]:
```
Iris_data.describe()
```

[ ]:
```
       sepal_length  sepal_width  petal_length  petal_width
count    150.000000   150.000000    150.000000   150.000000
mean       5.843333     3.054000      3.758667     1.198667
std        0.828066     0.433594      1.764420     0.763161
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
```

[ ]:
```
Iris_data.columns
```

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

[ ]:
```
Iris_data = Iris_data.rename(columns= {'species' : 'flower_type'})
```

[ ]:
```
Iris_data
```

[ ]:
```
     sepal_length  sepal_width  petal_length  petal_width     flower_type
0             5.1          3.5           1.4          0.2     Iris-setosa
1             4.9          3.0           1.4          0.2     Iris-setosa
2             4.7          3.2           1.3          0.2     Iris-setosa
3             4.6          3.1           1.5          0.2     Iris-setosa
4             5.0          3.6           1.4          0.2     Iris-setosa
..            …            …             …            …             …
145           6.7          3.0           5.2          2.3  Iris-virginica
146           6.3          2.5           5.0          1.9  Iris-virginica
147           6.5          3.0           5.2          2.0  Iris-virginica
148           6.2          3.4           5.4          2.3  Iris-virginica
149           5.9          3.0           5.1          1.8  Iris-virginica

[150 rows x 5 columns]
```

[ ]:
```
Iris_data.isna()
```

[ ]:
```
     sepal_length  sepal_width  petal_length  petal_width  flower_type
0           False        False         False        False        False
1           False        False         False        False        False
2           False        False         False        False        False
3           False        False         False        False        False
4           False        False         False        False        False
..            …            …             …            …            …
145         False        False         False        False        False
146         False        False         False        False        False
```

```
147         False        False         False        False        False
148         False        False         False        False        False
149         False        False         False        False        False

[150 rows x 5 columns]
```

[ ]:
```
Iris_data.isna().sum()
```

[ ]:
```
sepal_length    0
sepal_width     0
petal_length    0
petal_width     0
flower_type     0
dtype: int64
```

[ ]:
```
Iris_data.duplicated()
```

[ ]:
```
0      False
1      False
2      False
3      False
4      False
       …
145    False
146    False
147    False
148    False
149    False
Length: 150, dtype: bool
```

[ ]:
```
Iris_data[Iris_data.duplicated()]
```

[ ]:
```
     sepal_length  sepal_width  petal_length  petal_width     flower_type
34            4.9          3.1           1.5          0.1     Iris-setosa
37            4.9          3.1           1.5          0.1     Iris-setosa
142           5.8          2.7           5.1          1.9  Iris-virginica
```

[ ]:
```
Iris_data.query('sepal_width == 2.7 & sepal_length == 5.8 & petal_length == 5.
↪1')
```

[ ]:
```
     sepal_length  sepal_width  petal_length  petal_width     flower_type
101           5.8          2.7           5.1          1.9  Iris-virginica
142           5.8          2.7           5.1          1.9  Iris-virginica
```

[ ]:
```
Iris_data1 = ~Iris_data.duplicated(subset␣
↪=['sepal_length','sepal_width','petal_length','petal_width','flower_type'])
```

[ ]:
```
Iris_data1
```

[ ]:
```
0      True
1      True
2      True
3      True
4      True
       …
145    True
146    True
147    True
148    True
149    True
Length: 150, dtype: bool
```

[ ]:
```
Iris_data1 = ~Iris_data.duplicated(subset␣
↪=['sepal_length','sepal_width','petal_length','petal_width','flower_type'])
```

[ ]:
```
Iris_data1
```

[ ]:
```
0      True
1      True
2      True
3      True
4      True
       …
145    True
146    True
147    True
148    True
149    True
Length: 150, dtype: bool
```

[ ]:
```
Iris_data1 = Iris_data.loc[~Iris_data.duplicated(subset␣
↪=['sepal_length','sepal_width','petal_length','petal_width','flower_type'])]
```

[ ]:
```
Iris_data1.loc[Iris_data.duplicated()]
```

[ ]:
```
Empty DataFrame
Columns: [sepal_length, sepal_width, petal_length, petal_width, flower_type]
Index: []
```

[ ]:
```
Iris_data1.head(6)
```

[ ]:

	sepal_length	sepal_width	petal_length	petal_width	flower_type
0	5.1	3.5	1.4	0.2	Iris-setosa
1	4.9	3.0	1.4	0.2	Iris-setosa
2	4.7	3.2	1.3	0.2	Iris-setosa
3	4.6	3.1	1.5	0.2	Iris-setosa
4	5.0	3.6	1.4	0.2	Iris-setosa

```
5           5.4           3.9           1.7           0.4  Iris-setosa
```

[ ]:
```
Iris_data1.shape
```

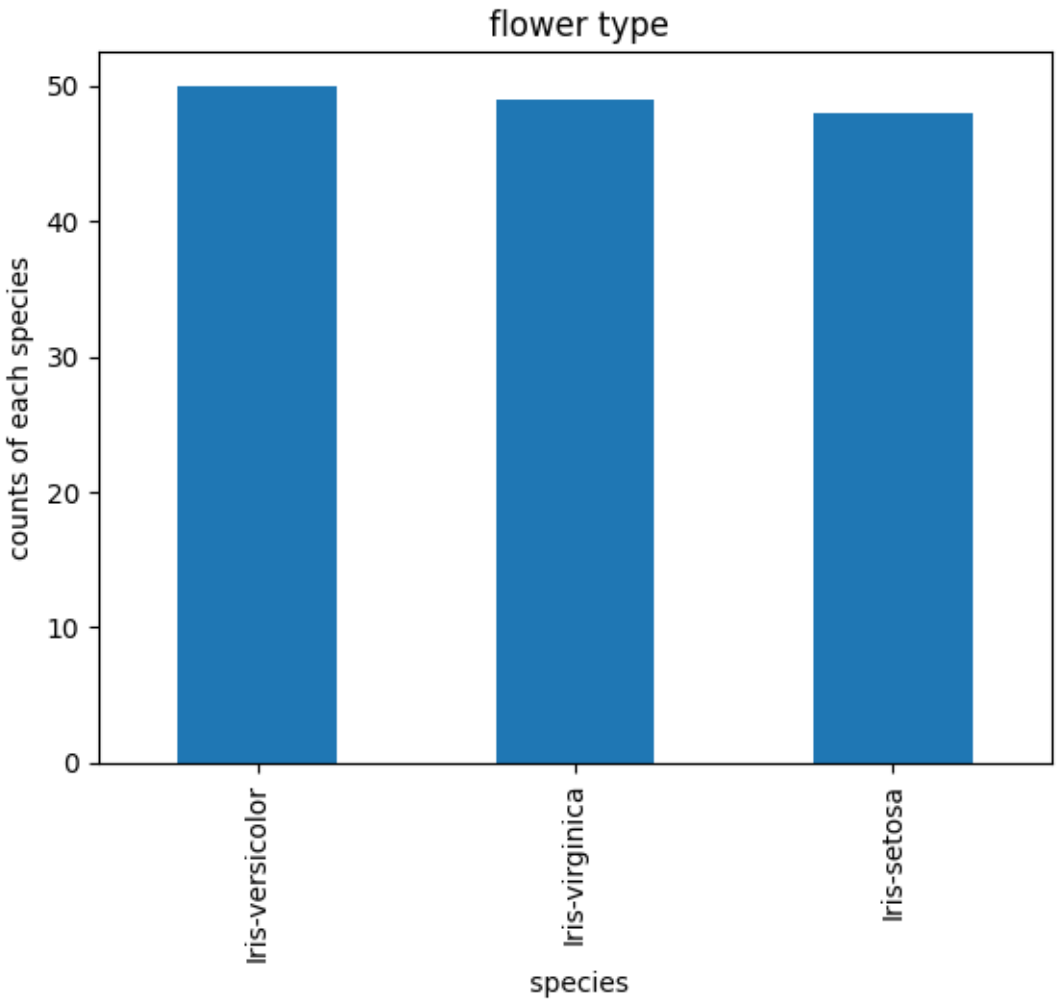
[ ]:
```
(147, 5)
```

[ ]:
```
# Feature Understanding
#histogram for each and every numeric variable

ax = Iris_data1['flower_type'].value_counts() \
    .head(6) \
    .plot(kind='bar',title=" flower type")

ax.set_xlabel(" species")
ax.set_ylabel("counts of each species")
```

[ ]:
```
Text(0, 0.5, 'counts of each species')
```

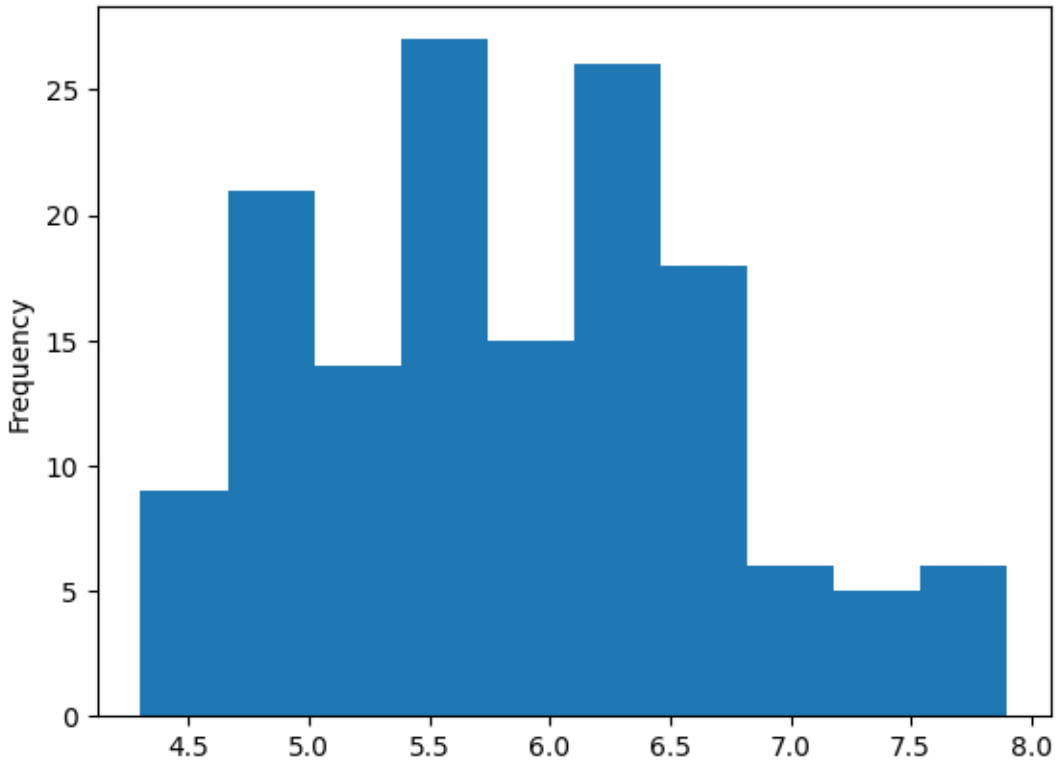


[ ]:
```
Iris_data1.columns
```

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

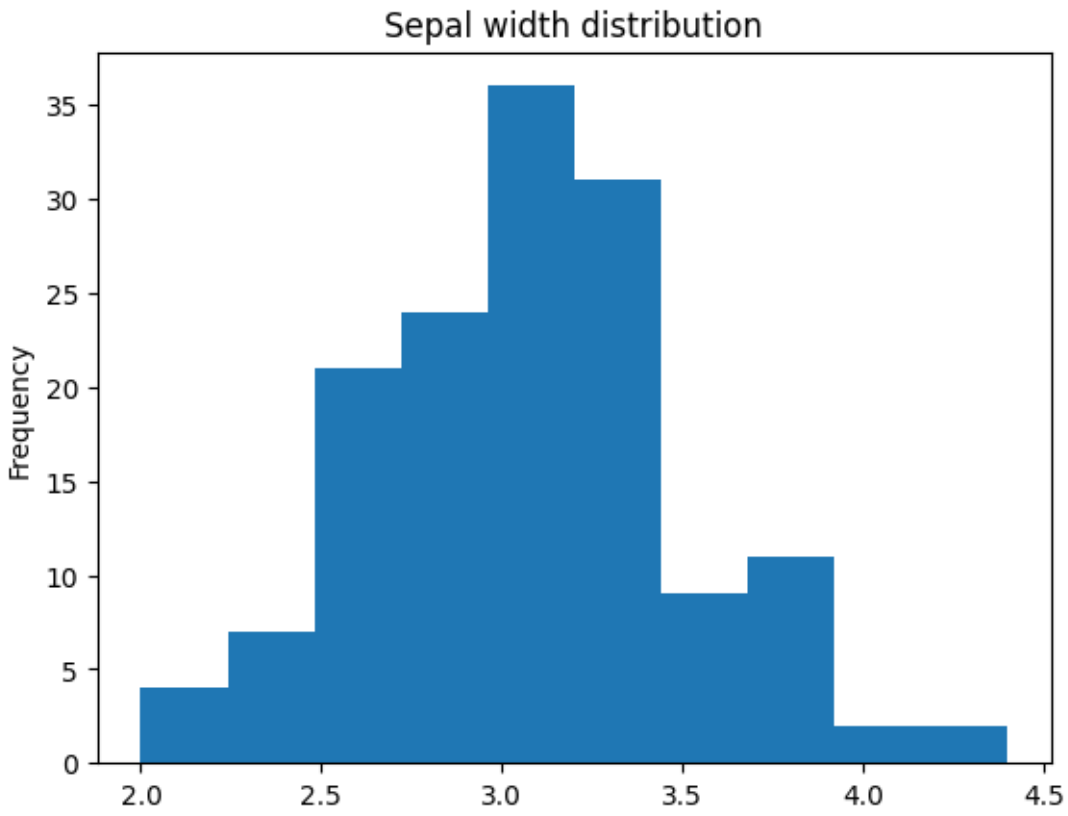
[ ]:
```
Iris_data1['sepal_length'].plot(kind='hist')
```

[ ]:
```
<Axes: ylabel='Frequency'>
```



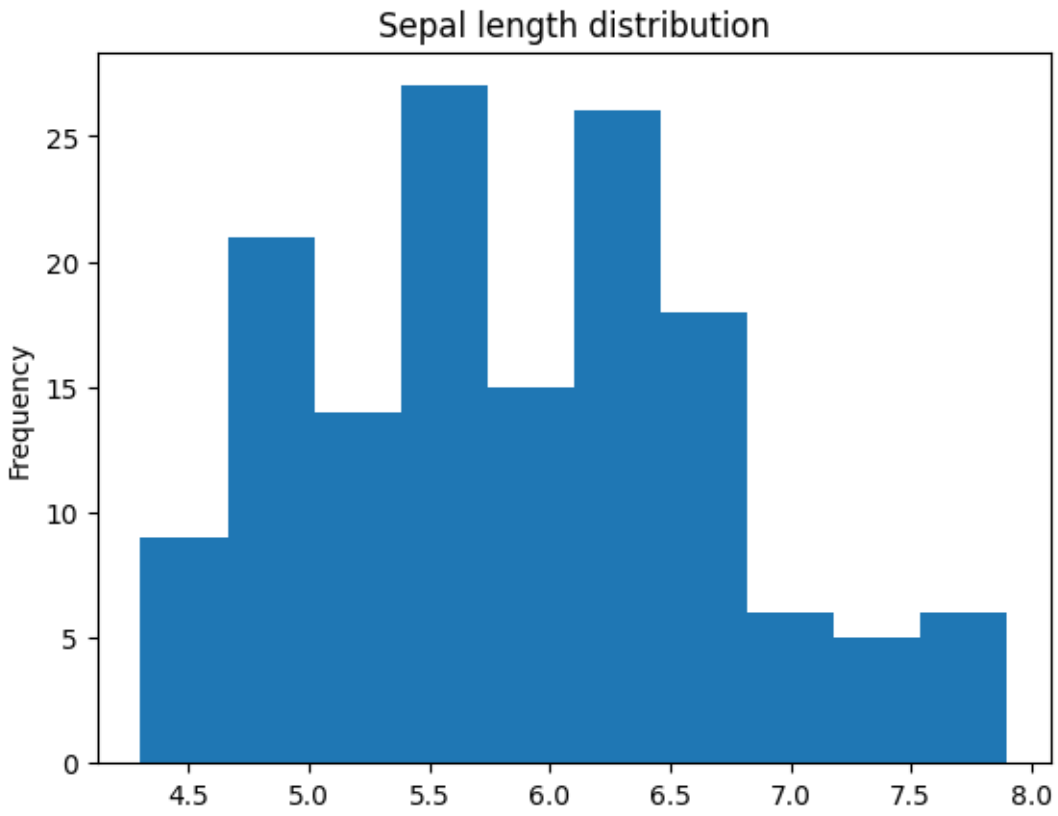
[ ]:
```
Iris_data1['sepal_width'].plot(kind='hist',title= "Sepal width distribution" )
```

[ ]:
```
<Axes: title={'center': 'Sepal width distribution'}, ylabel='Frequency'>
```



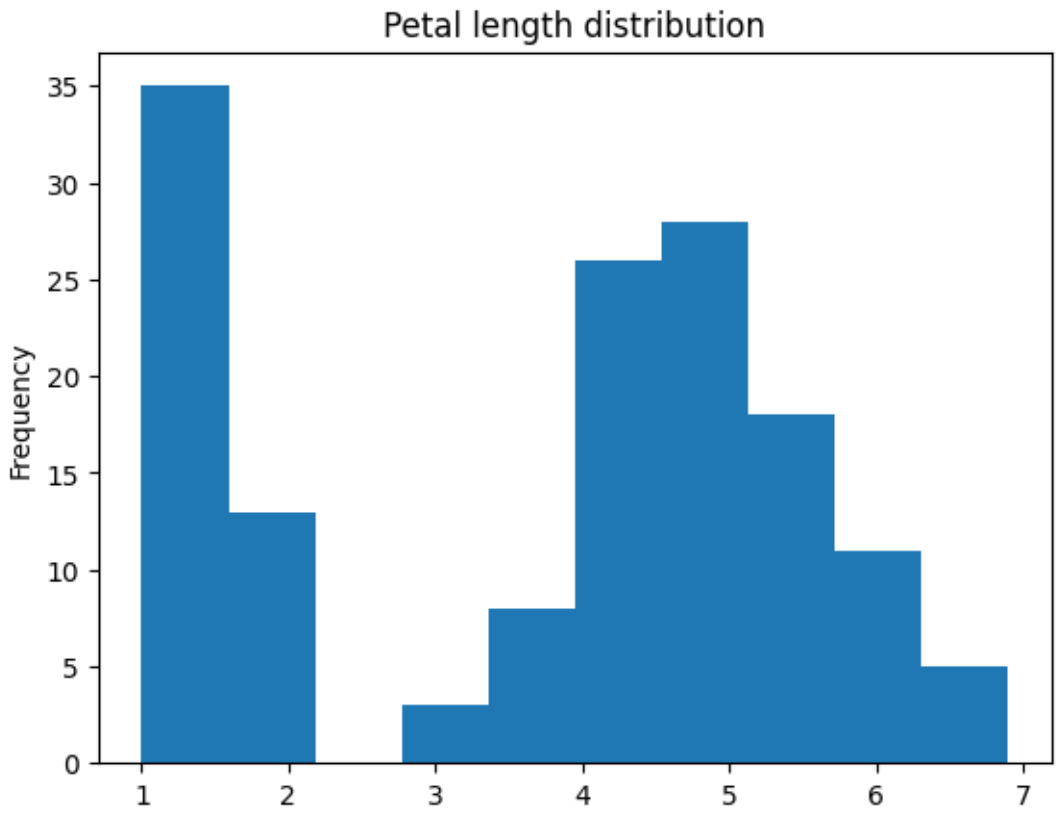
[ ]:
```
Iris_data1['sepal_length'].plot(kind='hist',title= "Sepal length distribution" )
```

[ ]:
```
<Axes: title={'center': 'Sepal length distribution'}, ylabel='Frequency'>
```



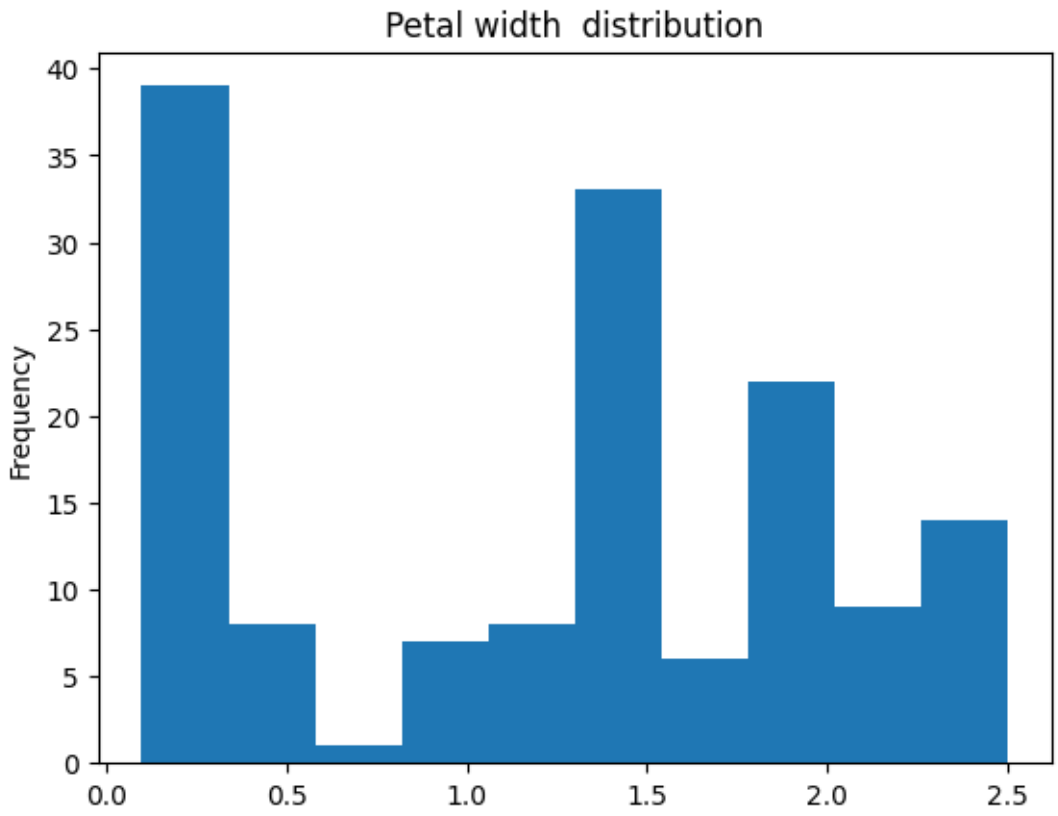
[ ]:
```
Iris_data1['petal_length'].plot(kind='hist',title= "Petal length distribution" )
```

[ ]:
```
<Axes: title={'center': 'Petal length distribution'}, ylabel='Frequency'>
```



[ ]:
```
Iris_data1['petal_width'].plot(kind='hist',title= "Petal width  distribution" )
```

[ ]:
```
<Axes: title={'center': 'Petal width  distribution'}, ylabel='Frequency'>
```



[ ]:
```
Iris_data1['flower_type']
```

[ ]:
```
0          Iris-setosa
1          Iris-setosa
2          Iris-setosa
3          Iris-setosa
4          Iris-setosa
              …
145    Iris-virginica
146    Iris-virginica
147    Iris-virginica
148    Iris-virginica
149    Iris-virginica
Name: flower_type, Length: 147, dtype: object
```

[ ]:
```
iris_setosa = Iris_data1.query('flower_type == "Iris-setosa" ')
iris_setosa.describe()

iris_versicolor = Iris_data1.query('flower_type== "Iris-versicolor"')
iris_versicolor.describe()
```

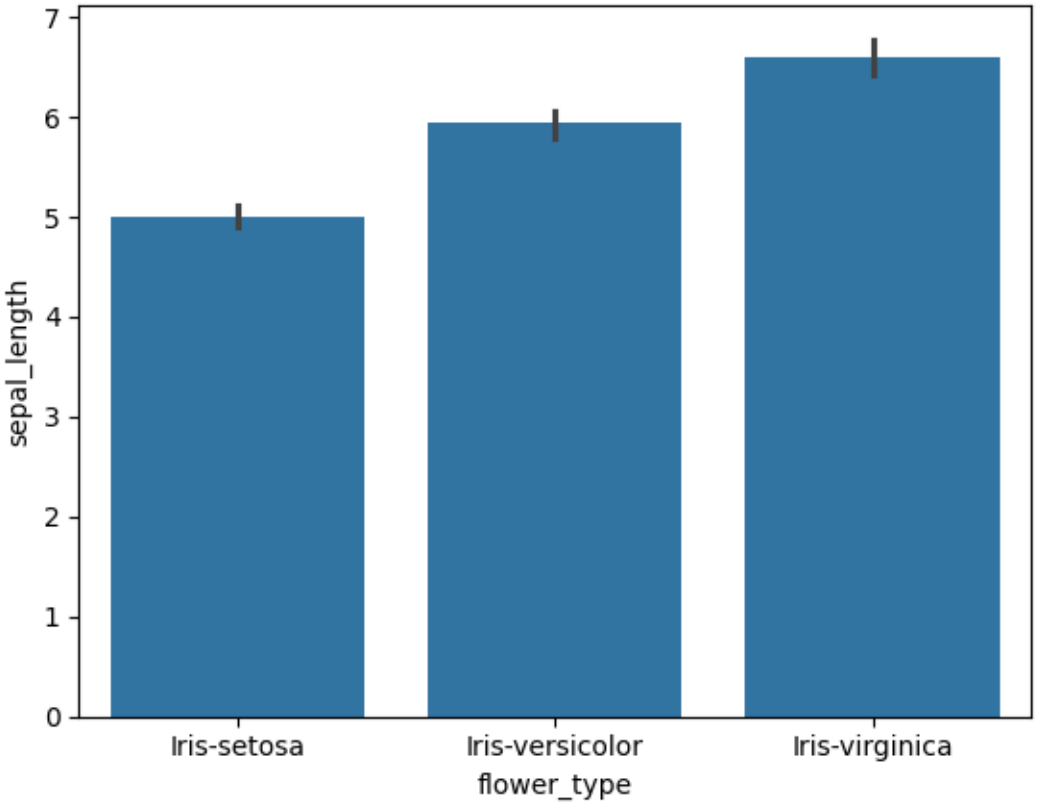
```
iris_virginica = Iris_data1.query('flower_type == "Iris-virginica"')
iris_virginica.describe()
```

[ ]:

	sepal_length	sepal_width	petal_length	petal_width
count	49.000000	49.000000	49.000000	49.000000
mean	6.604082	2.979592	5.561224	2.028571
std	0.632113	0.323380	0.553706	0.276887
min	4.900000	2.200000	4.500000	1.400000
25%	6.300000	2.800000	5.100000	1.800000
50%	6.500000	3.000000	5.600000	2.000000
75%	6.900000	3.200000	5.900000	2.300000
max	7.900000	3.800000	6.900000	2.500000

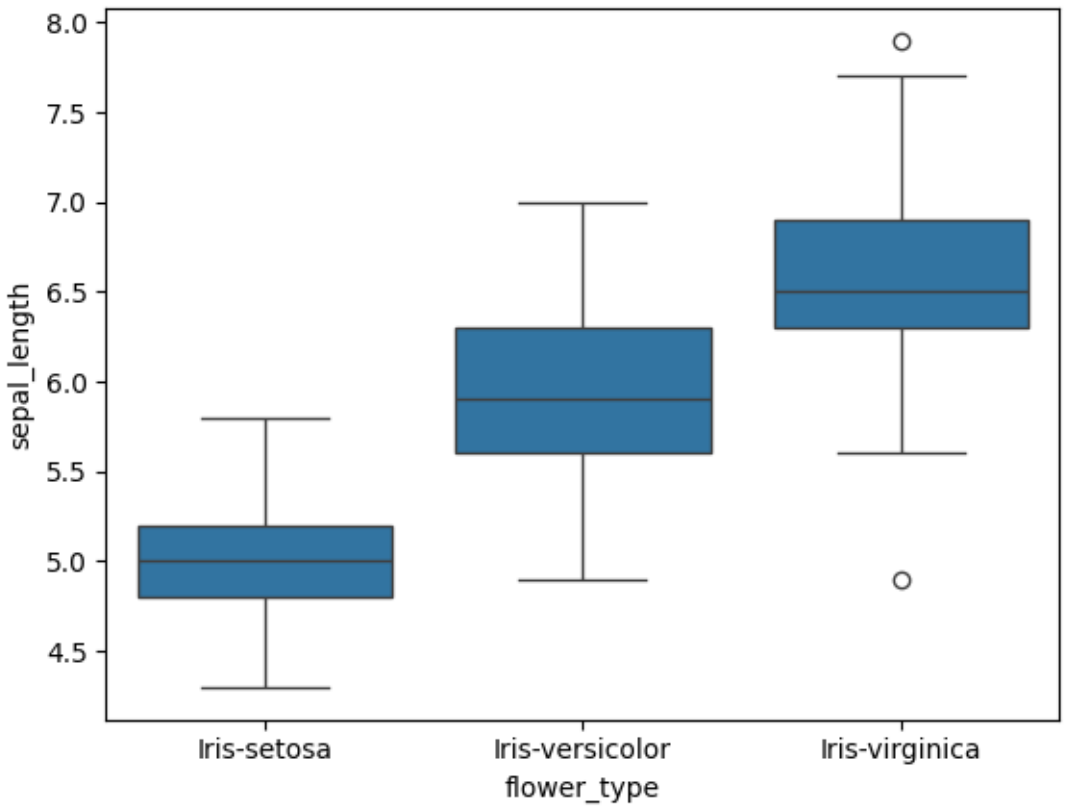
[ ]:
```
sns.barplot(x= 'flower_type', y= "sepal_length",data = Iris_data1)
```

[ ]:
```
<Axes: xlabel='flower_type', ylabel='sepal_length'>
```



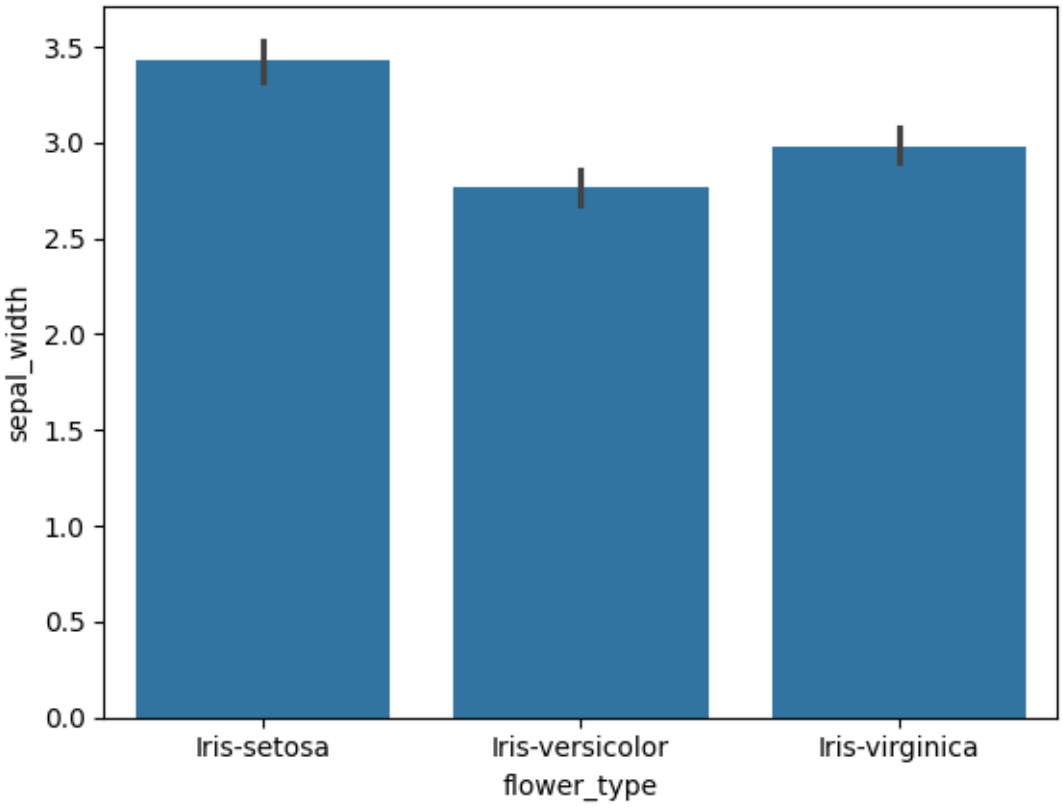
[ ]:
```
sns.boxplot(x= 'flower_type',y= "sepal_length",data = Iris_data1)
```

[ ]:
```
<Axes: xlabel='flower_type', ylabel='sepal_length'>
```



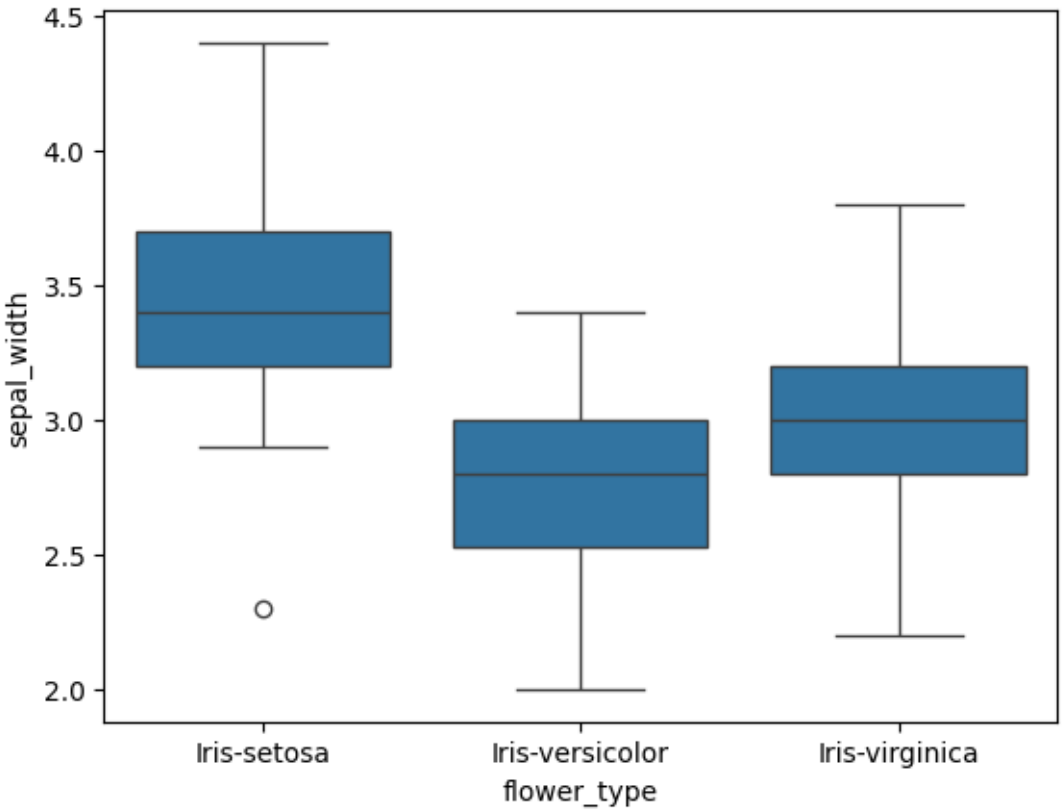
[ ]:
```
sns.barplot(x= 'flower_type', y= "sepal_width",data = Iris_data1)
```

[ ]: <Axes: xlabel='flower_type', ylabel='sepal_width'>



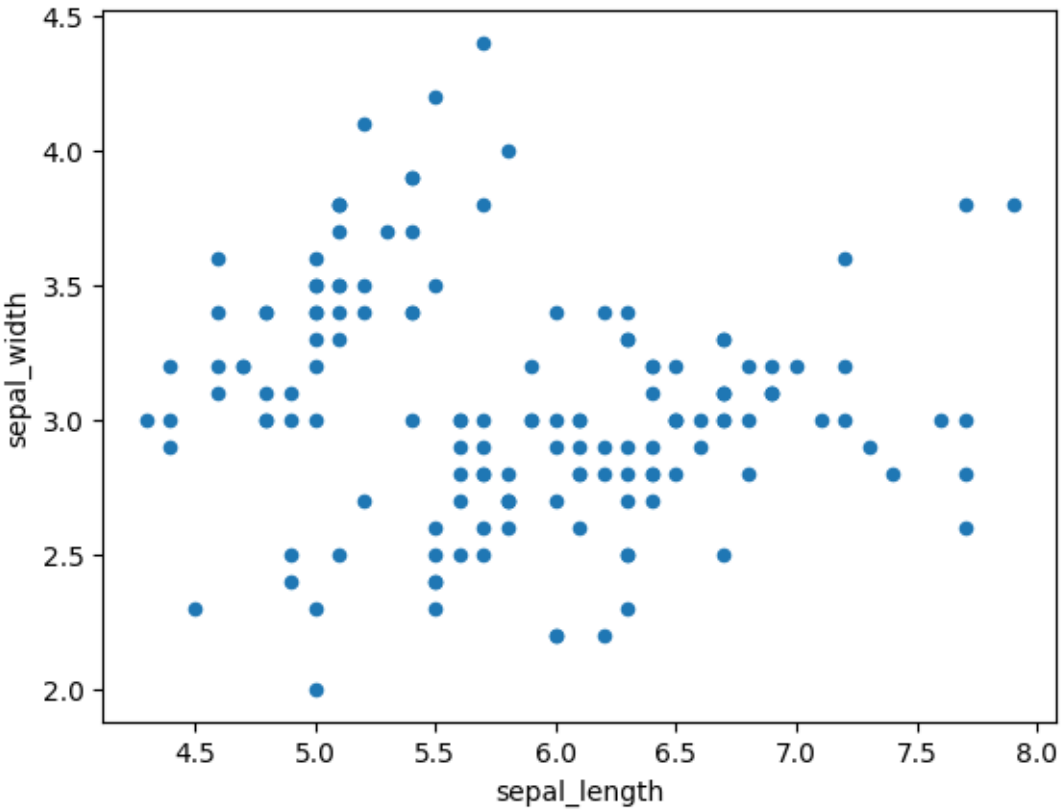
[ ]:
```
sns.boxplot(x= 'flower_type',y= "sepal_width",data = Iris_data1)
```

[ ]:
```
<Axes: xlabel='flower_type', ylabel='sepal_width'>
```



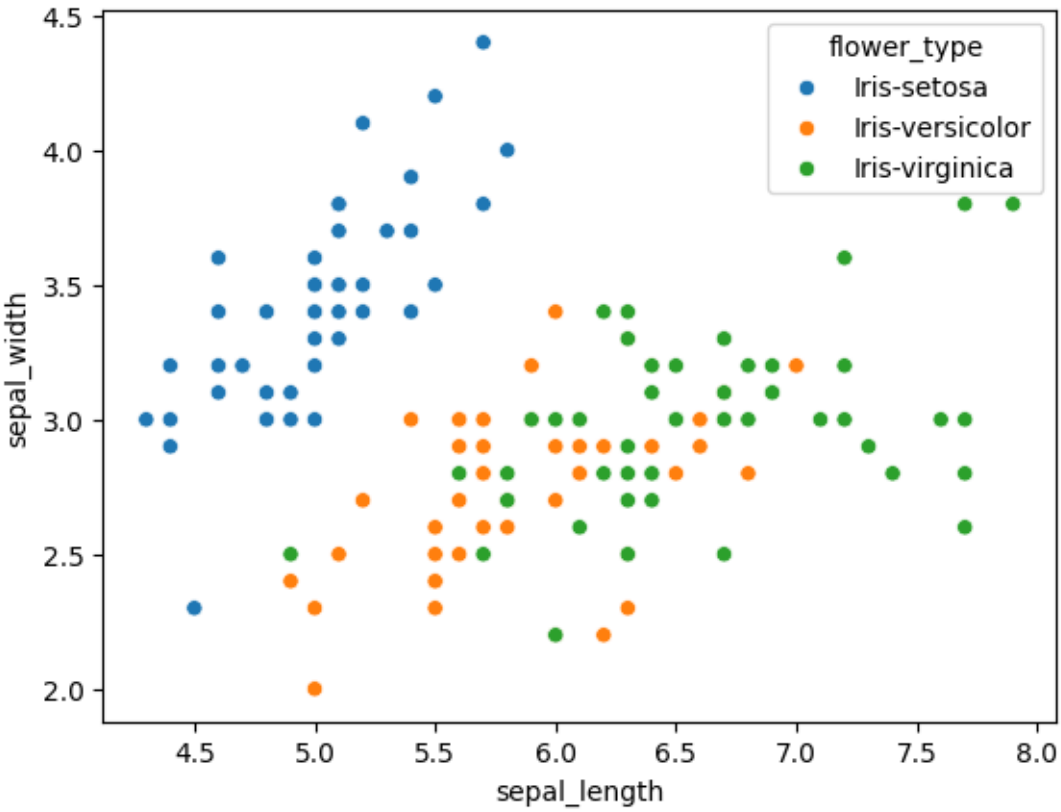
[ ]:
```
Iris_data1.plot(kind= 'scatter',
                                x="sepal_length",y="sepal_width")
```

[ ]: <Axes: xlabel='sepal_length', ylabel='sepal_width'>



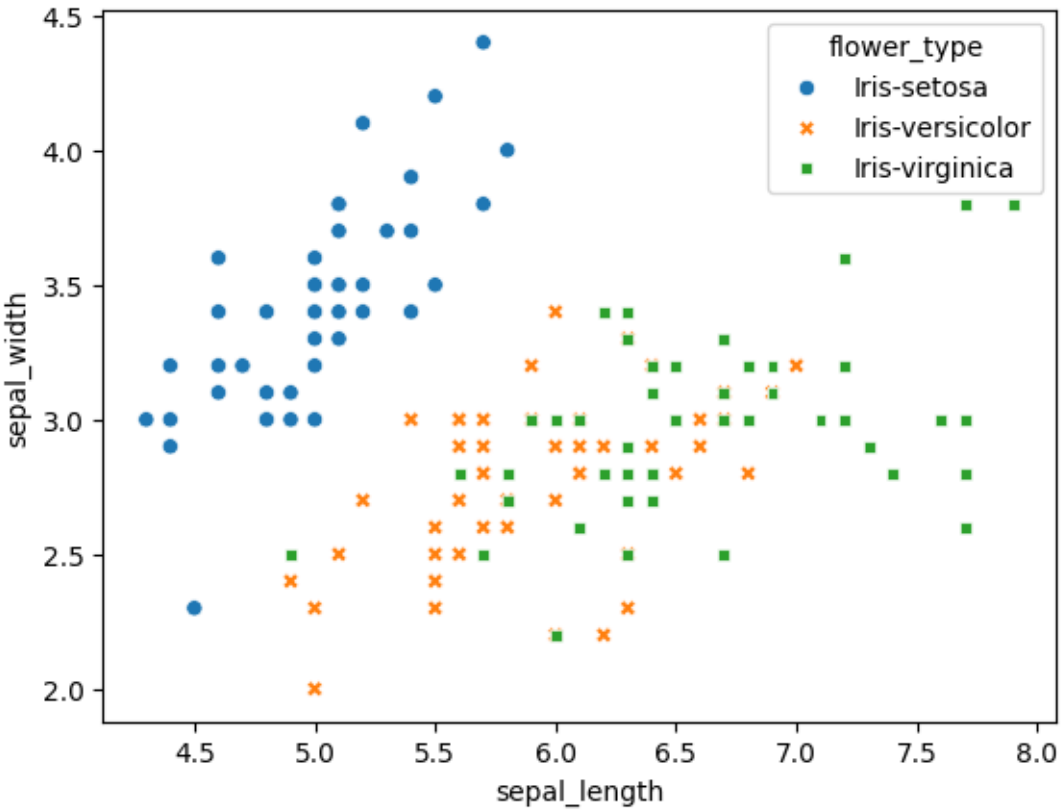
[ ]:
```
sns.scatterplot(x='sepal_length',y="sepal_width",hue='flower_type',data =␣
↪Iris_data1)
```

[ ]:
```
<Axes: xlabel='sepal_length', ylabel='sepal_width'>
```



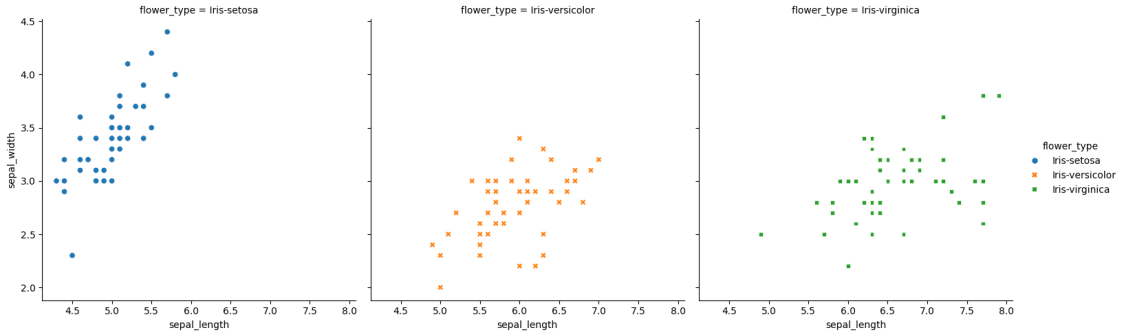
[ ]:
```
sns.scatterplot(x='sepal_length',y="sepal_width",hue='flower_type',data =␣
↪Iris_data1,style='flower_type')
```

[ ]:
```
<Axes: xlabel='sepal_length', ylabel='sepal_width'>
```



[ ]:
```
sns.
↪relplot(x='sepal_length',y="sepal_width",hue='flower_type',col='flower_type',data␣
↪= Iris_data1,style='flower_type')
```

[ ]:
```
<seaborn.axisgrid.FacetGrid at 0x1f33435f4d0>
```

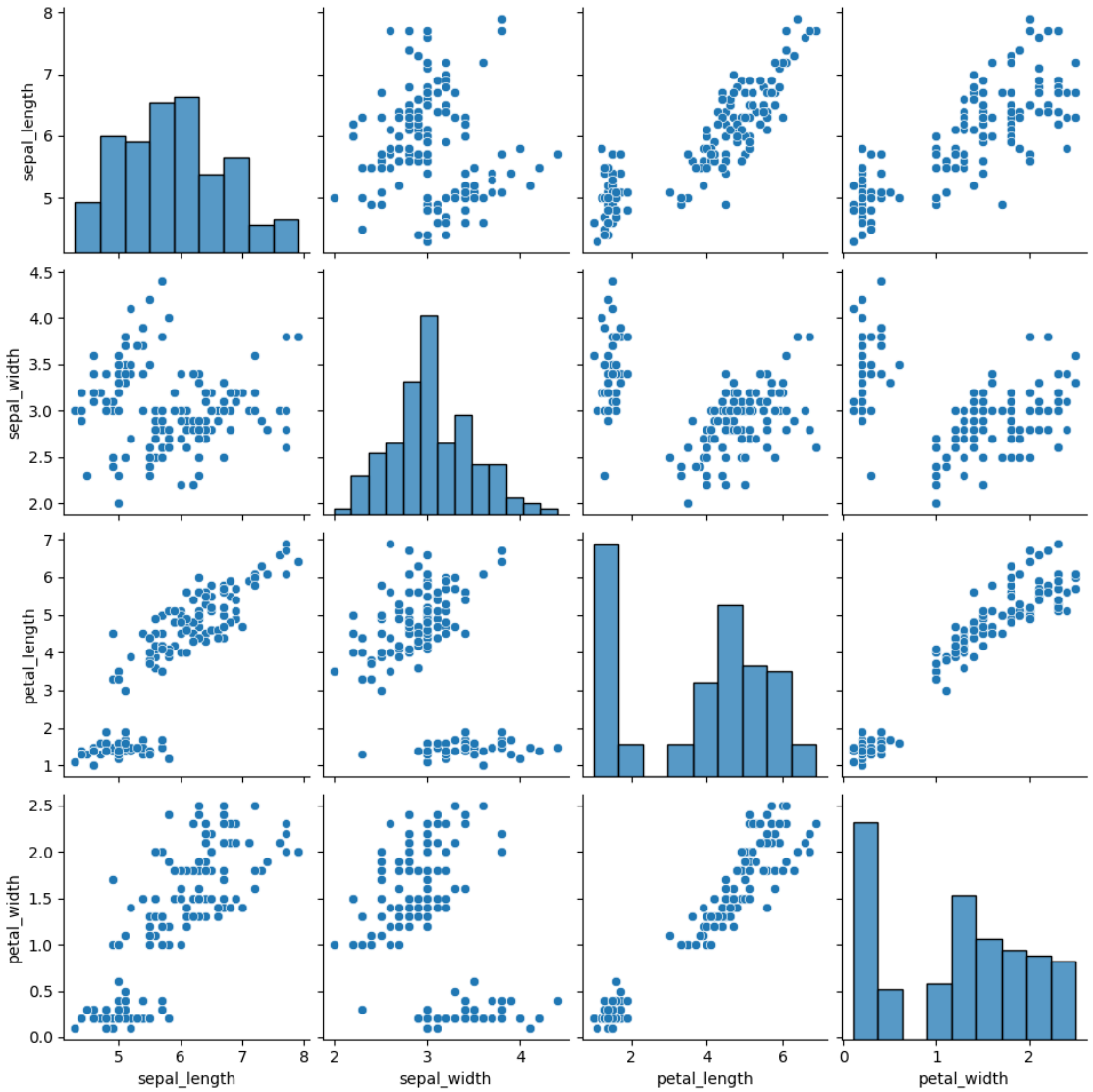


[ ]:
```
Iris_data1.columns
```

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

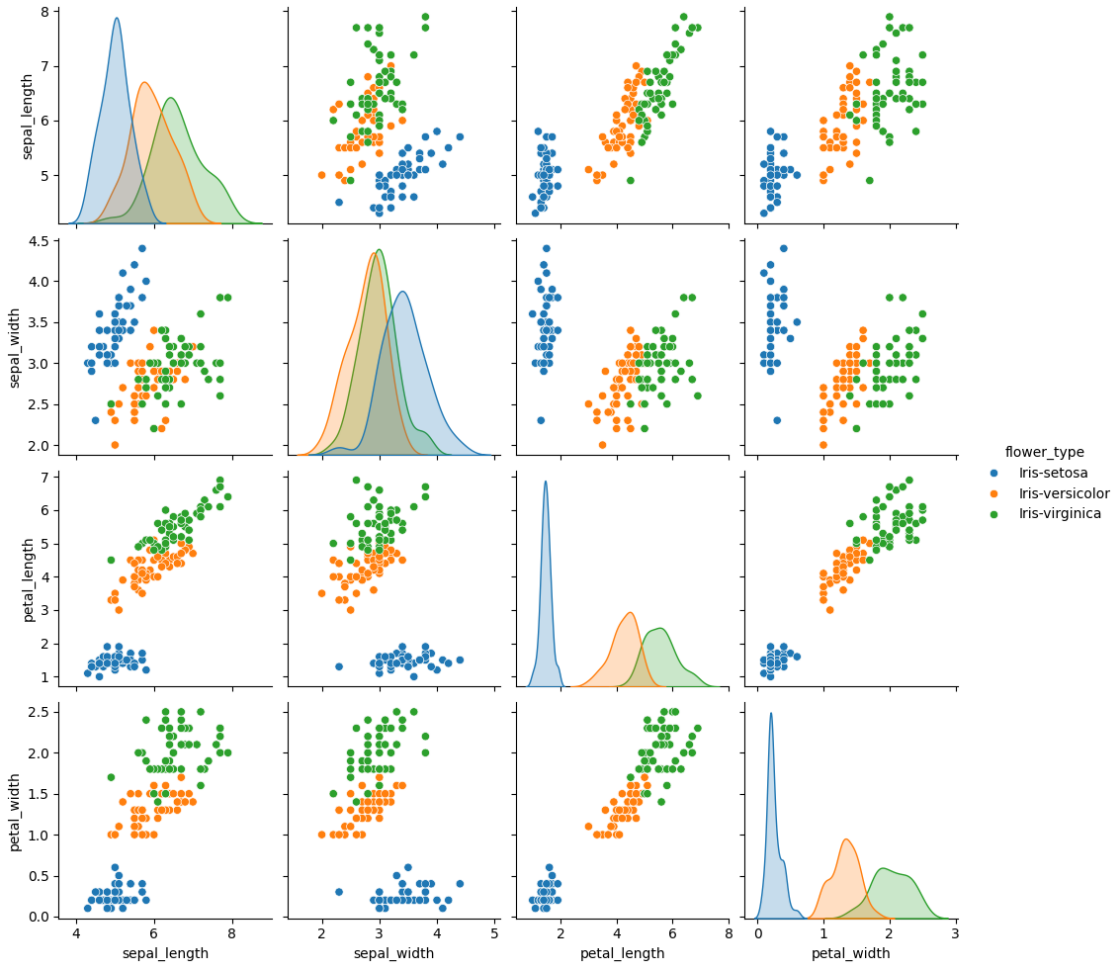
[ ]:
```
sns.pairplot(Iris_data1,vars= ['sepal_length', 'sepal_width', 'petal_length',␣
↪'petal_width'] )
```

[ ]:
```
<seaborn.axisgrid.PairGrid at 0x1f33435f770>
```



[ ]:
```
sns.pairplot(Iris_data1,vars= ['sepal_length', 'sepal_width', 'petal_length',␣
↪'petal_width'],hue="flower_type")
```

[ ]:
```
<seaborn.axisgrid.PairGrid at 0x1f33c540190>
```



[ ]:
```
# get the plot correlation matrix of the variables
Iris_data1[['sepal_length', 'sepal_width', 'petal_length', 'petal_width']].
↪dropna().corr()
```

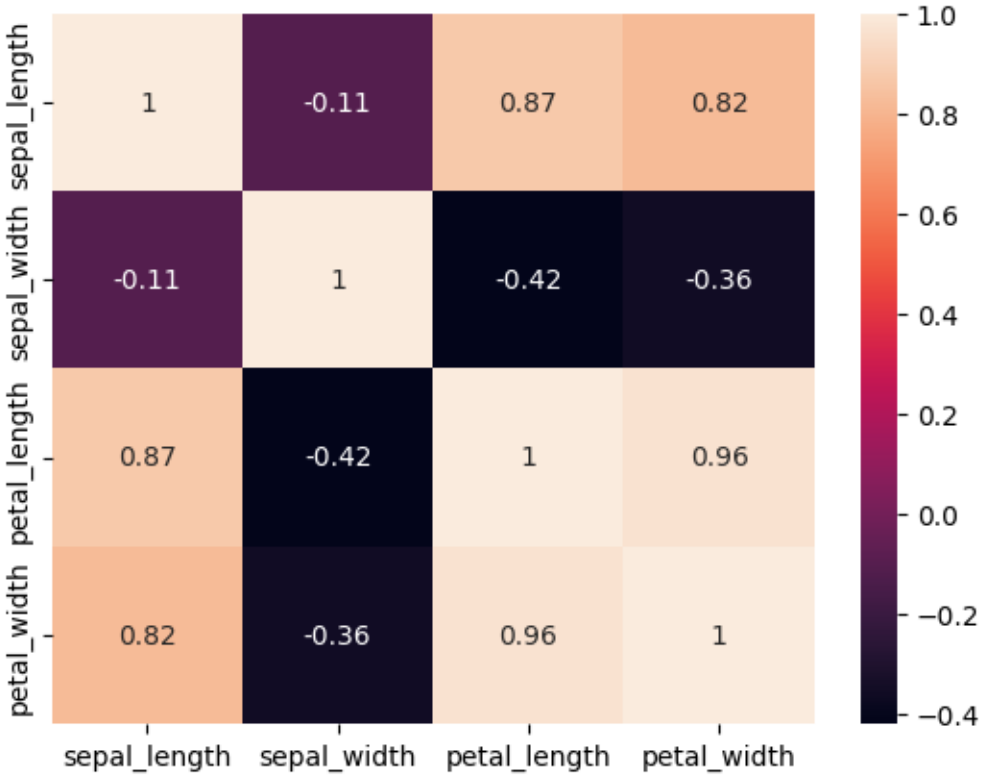
[ ]:

	sepal_length	sepal_width	petal_length	petal_width
sepal_length	1.000000	-0.109321	0.871305	0.817058
sepal_width	-0.109321	1.000000	-0.421057	-0.356376
petal_length	0.871305	-0.421057	1.000000	0.961883
petal_width	0.817058	-0.356376	0.961883	1.000000

[ ]:
```
corr_df = Iris_data1[['sepal_length', 'sepal_width',
                      'petal_length', 'petal_width']].dropna().corr()
```

[ ]:
```
sns.heatmap(corr_df,annot=True)
```

[ ]: <Axes: >


sepal_length
sepal_width
petal_length
petal_width
1
-0.11
0.87
0.82
-0.11
1
-0.42
-0.36
0.87
-0.42
1
0.96
0.82
-0.36
0.96
1
sepal_length
sepal_width
petal_length
petal_width
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4