iris_dataVisualization

May 11, 2022

0.0.1 Data Visualization III

Download the Iris flower dataset or any other dataset into a DataFrame. (e.g., https://archive.ics.uci.edu/ml/datasets/Iris). Scan the dataset and give the inference as: - List down the features and their types (e.g., numeric, nominal) available in the dataset. - Create a histogram for each feature in the dataset to illustrate the feature distributions. - Create a boxplot for each feature in the dataset. - Compare distributions and identify outliers.

[5]:
```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib
import matplotlib.pyplot as plt
%matplotlib inline
url="Iris.csv"
df = pd.read_csv(url)
df.head(10)
```

[5]:

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
0	1	5.1	3.5	1.4	0.2	Iris-setosa
1	2	4.9	3.0	1.4	0.2	Iris-setosa
2	3	4.7	3.2	1.3	0.2	Iris-setosa
3	4	4.6	3.1	1.5	0.2	Iris-setosa
4	5	5.0	3.6	1.4	0.2	Iris-setosa
5	6	5.4	3.9	1.7	0.4	Iris-setosa
6	7	4.6	3.4	1.4	0.3	Iris-setosa
7	8	5.0	3.4	1.5	0.2	Iris-setosa
8	9	4.4	2.9	1.4	0.2	Iris-setosa
9	10	4.9	3.1	1.5	0.1	Iris-setosa

[67]:
```
df.describe()
```

[67]:

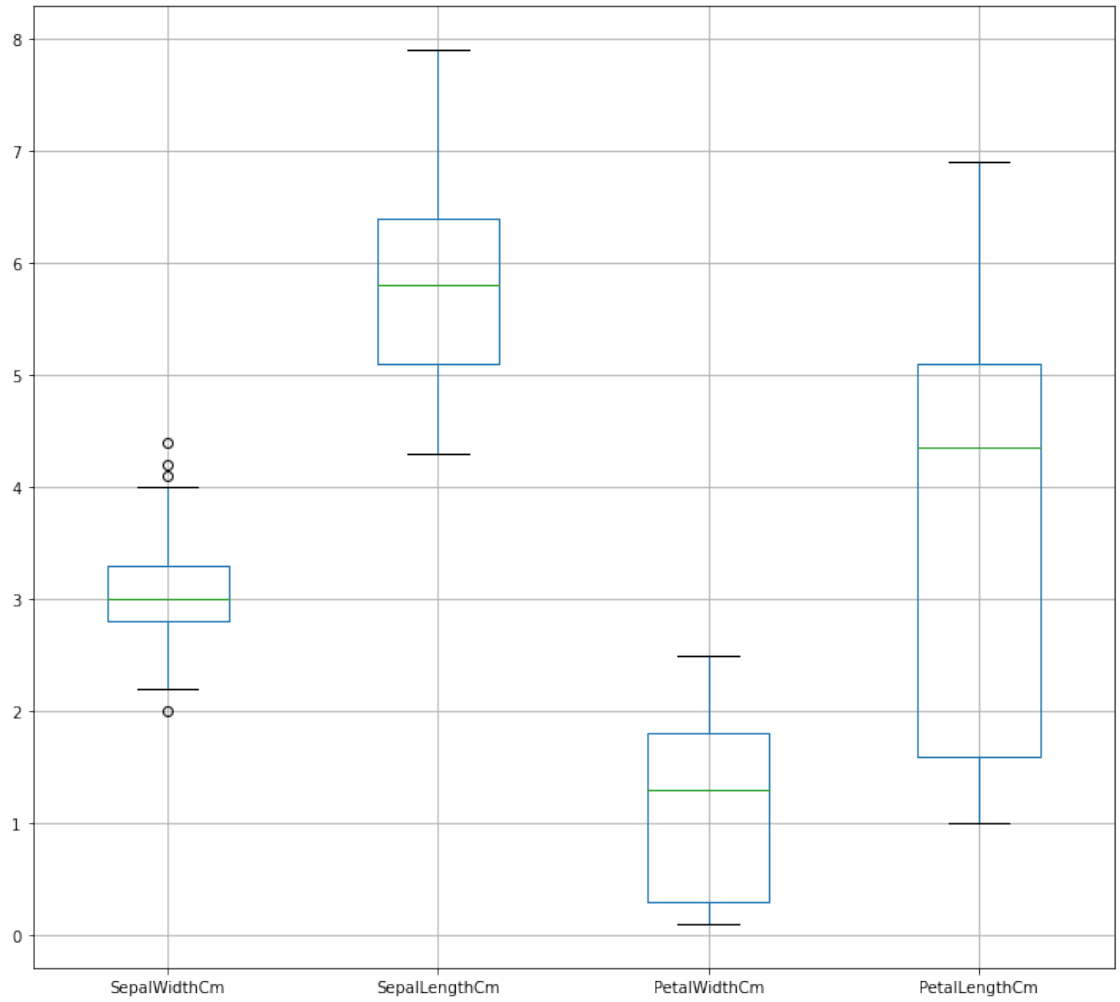
	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm
count	150.000000	150.000000	150.000000	150.000000	150.000000
mean	75.500000	5.843333	3.054000	3.758667	1.198667
std	43.445368	0.828066	0.433594	1.764420	0.763161
min	1.000000	4.300000	2.000000	1.000000	0.100000
25%	38.250000	5.100000	2.800000	1.600000	0.300000

```
50%     75.500000       5.800000       3.000000       4.350000       1.300000
75%    112.750000       6.400000       3.300000       5.100000       1.800000
max    150.000000       7.900000       4.400000       6.900000       2.500000
```

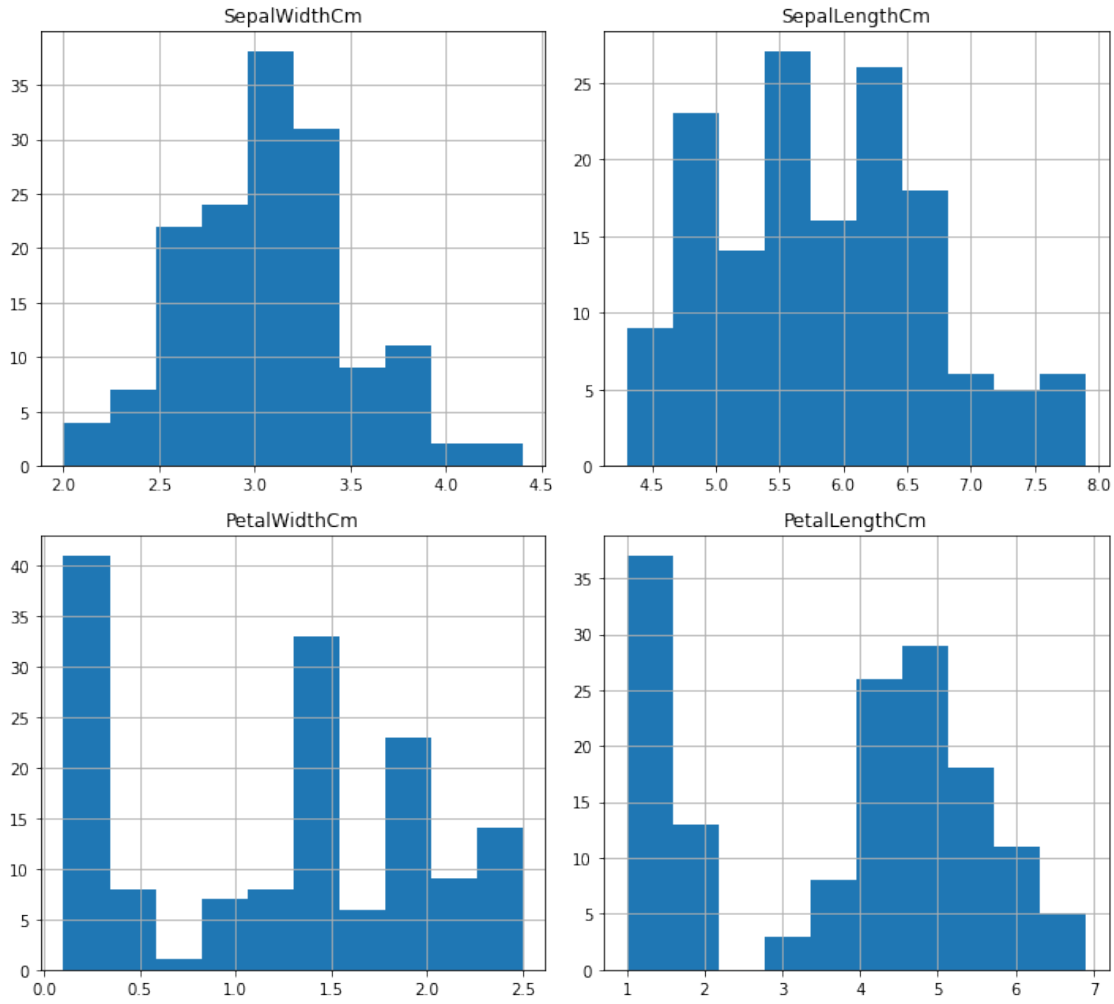
[6]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 150 entries, 0 to 149
Data columns (total 6 columns):
 #   Column         Non-Null Count  Dtype
---  ------         --------------  -----
 0   Id             150 non-null    int64
 1   SepalLengthCm  150 non-null    float64
 2   SepalWidthCm   150 non-null    float64
 3   PetalLengthCm  150 non-null    float64
 4   PetalWidthCm   150 non-null    float64
 5   Species        150 non-null    object
dtypes: float64(4), int64(1), object(1)
memory usage: 7.2+ KB
```

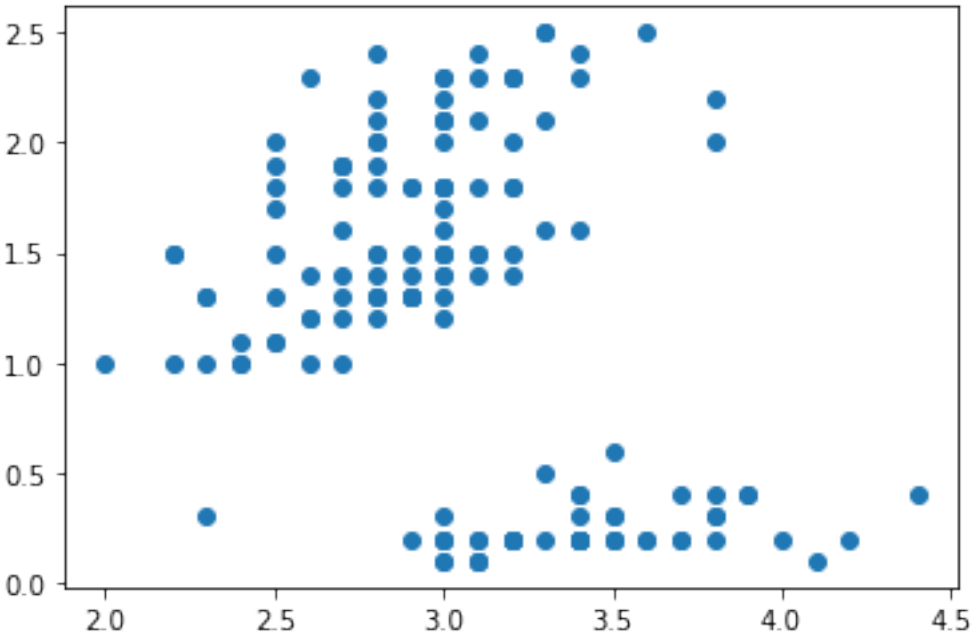
[66]:
```
#boxplot
df.
 ↪boxplot(column=["SepalWidthCm","SepalLengthCm","PetalWidthCm","PetalLengthCm"],figsize=(10,
plt.tight_layout()
plt.show()
```



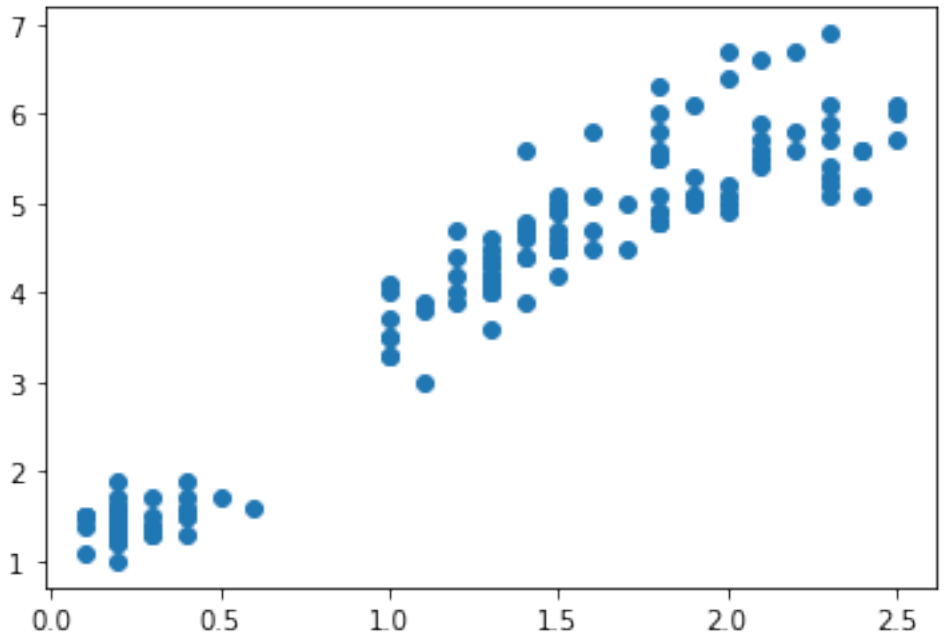
[63]:
```
#histogram
df.
 ↪hist(column=["SepalWidthCm","SepalLengthCm","PetalWidthCm","PetalLengthCm"],figsize=(10,9))
plt.tight_layout()
plt.show()
```



[76]:
```
plt.scatter(df["SepalWidthCm"],df["PetalWidthCm"])
plt.show()
```



[73]:
```
plt.scatter(df["PetalWidthCm"],df["PetalLengthCm"])
plt.show()
```



[77]:
```
plt.scatter(df["SepalLengthCm"],df["PetalWidthCm"])
plt.show()
```