

```
In [13]: pip install openpyxl
```

Requirement already satisfied: openpyxl in /srv/conda/envs/notebook/lib/python3.7/site-packages (3.0.7)
Requirement already satisfied: et-xmlfile in /srv/conda/envs/notebook/lib/python3.7/site-packages (from openpyxl) (1.1.0)
Note: you may need to restart the kernel to use updated packages.

```
In [15]: import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

```
In [5]: data=pd.read_excel("iris_data.xlsx")
```

```
In [6]: data
```

```
Out[6]:
```

	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 × 5 columns

```
In [7]: df=pd.DataFrame(data)
```

```
In [8]: 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]

```
In [9]: df.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
```

```
In [10]: df.describe()
```

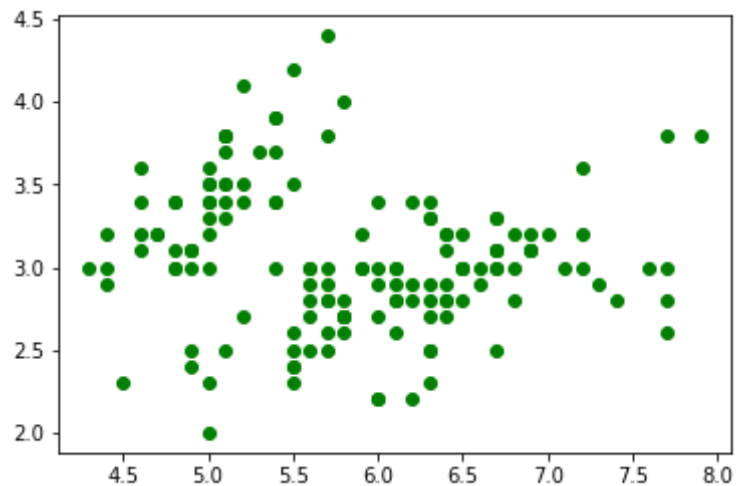
```
Out[10]:
```

	sepal_length	sepal_width	petal_length	petal_width
--	--------------	-------------	--------------	-------------

	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

```
In [22]: plt.plot(df.sepal_length,df.sepal_width,ls='',marker='o',color='green',label='sepal')
```

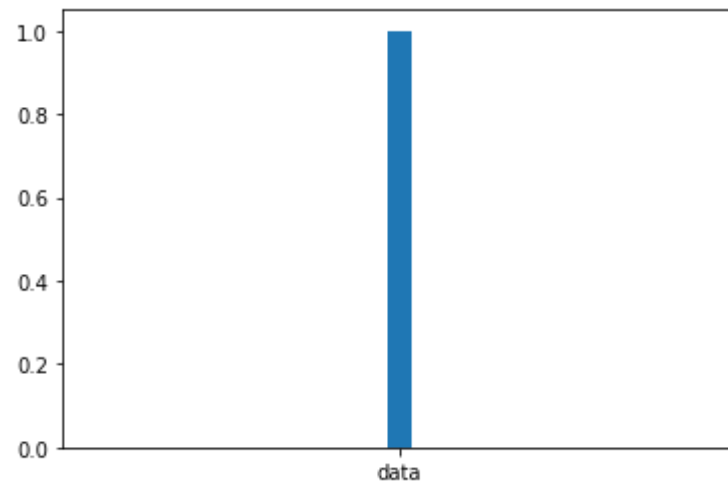
```
Out[22]: [<matplotlib.lines.Line2D at 0x7f27f4cd5610>]
```



```
In [20]: plt.hist('data',bins=25)
```

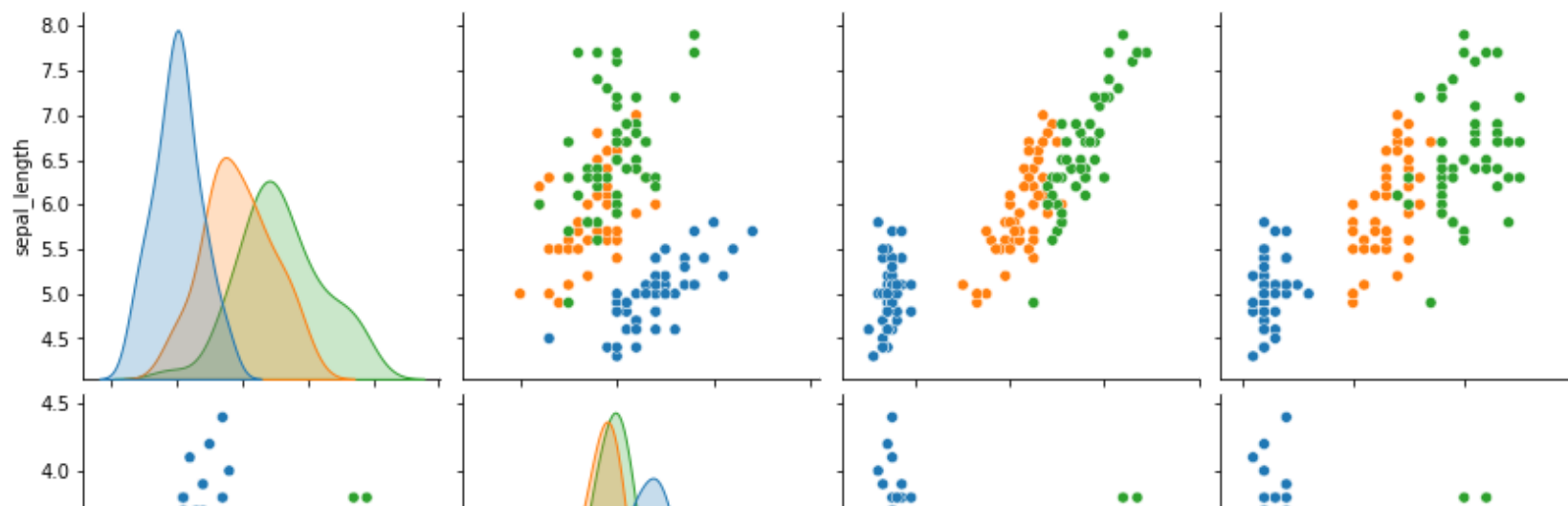
```
Out[20]: (array([0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 1., 0., 0., 0., 0.,
```

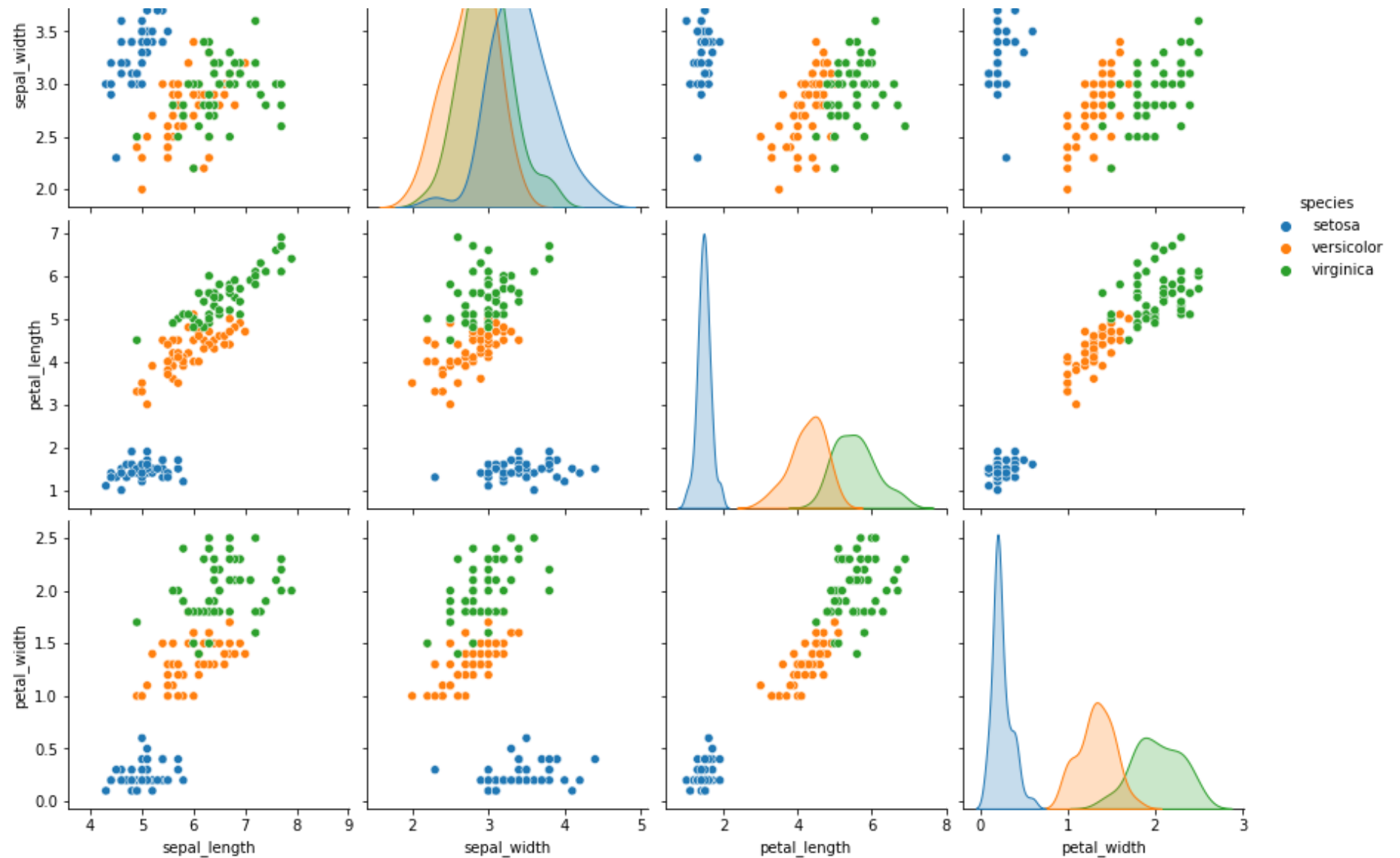
```
0., 0., 0., 0., 0., 0., 0., 0.] ),
array([-0.5 , -0.46, -0.42, -0.38, -0.34, -0.3 , -0.26, -0.22, -0.18,
       -0.14, -0.1 , -0.06, -0.02,  0.02,  0.06,  0.1 ,  0.14,  0.18,
        0.22,  0.26,  0.3 ,  0.34,  0.38,  0.42,  0.46,  0.5 ] ),
<a list of 25 Patch objects>)
```



```
In [24]: sns.pairplot(data,hue='species',height=3)
```

```
Out[24]: <seaborn.axisgrid.PairGrid at 0x7f27b7a37150>
```





```
In [27]: df.head()
```

```
Out[27]:
```

sepal_length	sepal_width	petal_length	petal_width	species
--------------	-------------	--------------	-------------	---------

	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

In [28]: `df.tail()`

Out[28]:

	sepal_length	sepal_width	petal_length	petal_width	species
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

In [29]: `len(data)`

Out[29]: 150

In [31]: `data.shape[0]`

Out[31]: 150

In [32]: `data.shape`

Out[32]: (150, 5)

```
In [33]: data['species']=data.species.str.replace('Iris',' ')
```

```
In [34]: data.head()
```

```
Out[34]:
```

	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

```
In [35]: data.tail()
```

```
Out[35]:
```

	sepal_length	sepal_width	petal_length	petal_width	species
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

```
In [37]: print(df.value_counts().sum())
```

150

```
In [38]: print(df.value_counts())
```

sepal_length	sepal_width	petal_length	petal_width	species	
4.9	3.1	1.5	0.1	setosa	3
5.8	2.7	5.1	1.9	virginica	2
	4.0	1.2	0.2	setosa	1

```

5.9      3.0      4.2      1.5      versicolor      1
6.2      3.4      5.4      2.3      virginica        1
..
5.5      2.3      4.0      1.3      versicolor      1
      2.4      3.7      1.0      versicolor      1
      3.8      1.1      versicolor      1
      2.5      4.0      1.3      versicolor      1
7.9      3.8      6.4      2.0      virginica        1
Length: 147, dtype: int64

```

In [39]: `data.value_counts()`

```

Out[39]: sepal_length  sepal_width  petal_length  petal_width  species
4.9      3.1      1.5      0.1      setosa        3
5.8      2.7      5.1      1.9      virginica      2
      4.0      1.2      0.2      setosa        1
5.9      3.0      4.2      1.5      versicolor    1
6.2      3.4      5.4      2.3      virginica      1
..
5.5      2.3      4.0      1.3      versicolor    1
      2.4      3.7      1.0      versicolor    1
      3.8      1.1      versicolor    1
      2.5      4.0      1.3      versicolor    1
7.9      3.8      6.4      2.0      virginica      1
Length: 147, dtype: int64

```

In [40]: `data.describe()`

```

Out[40]:
   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

```


	sepal_length	sepal_width	petal_length	petal_width
max	7.900000	4.400000	6.900000	2.500000

```
In [41]: data.groupby('species').median()
```

```
Out[41]:
```

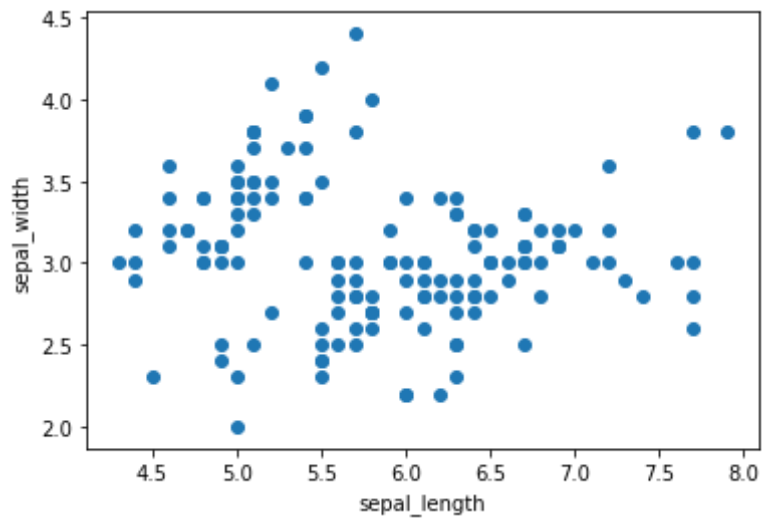
	sepal_length	sepal_width	petal_length	petal_width
species				
setosa	5.0	3.4	1.50	0.2
versicolor	5.9	2.8	4.35	1.3
virginica	6.5	3.0	5.55	2.0

```
In [42]: data.groupby('species').mean()
```

```
Out[42]:
```

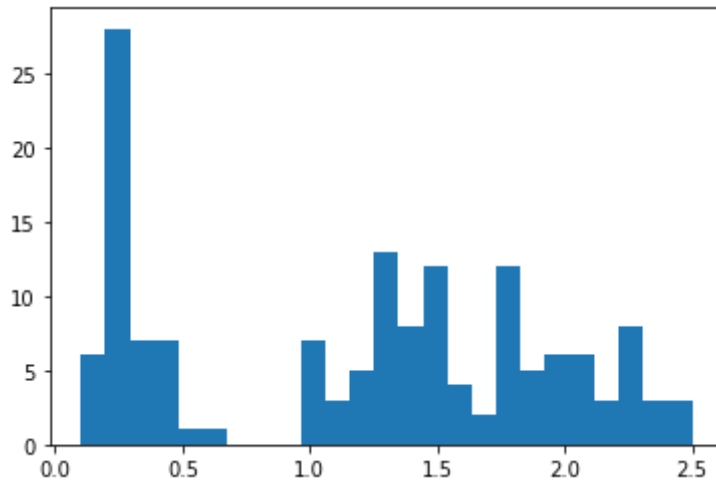
	sepal_length	sepal_width	petal_length	petal_width
species				
setosa	5.006	3.418	1.464	0.244
versicolor	5.936	2.770	4.260	1.326
virginica	6.588	2.974	5.552	2.026

```
In [51]: fig,ax=plt.subplots()
ax.scatter(data.sepal_length,data.sepal_width)
ax.set(xlabel='sepal_length',ylabel='sepal_width')
plt.show()
```



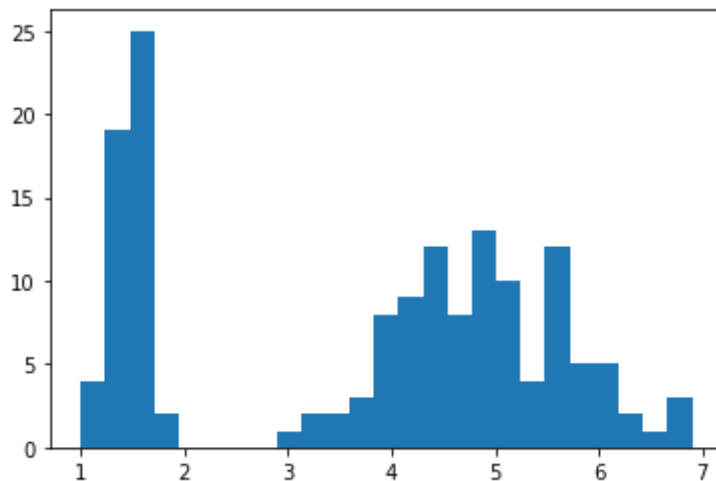
```
In [53]: fig,ax=plt.subplots()  
         ax.hist(data.petal_width,bins=25)
```

```
Out[53]: (array([ 6., 28.,  7.,  7.,  1.,  1.,  0.,  0.,  0.,  7.,  3.,  5., 13.,  
                  8., 12.,  4.,  2., 12.,  5.,  6.,  6.,  3.,  8.,  3.,  3.]),  
         array([0.1  , 0.196, 0.292, 0.388, 0.484, 0.58  , 0.676, 0.772, 0.868,  
                0.964, 1.06  , 1.156, 1.252, 1.348, 1.444, 1.54  , 1.636, 1.732,  
                1.828, 1.924, 2.02  , 2.116, 2.212, 2.308, 2.404, 2.5  ]),  
         <a list of 25 Patch objects>)
```

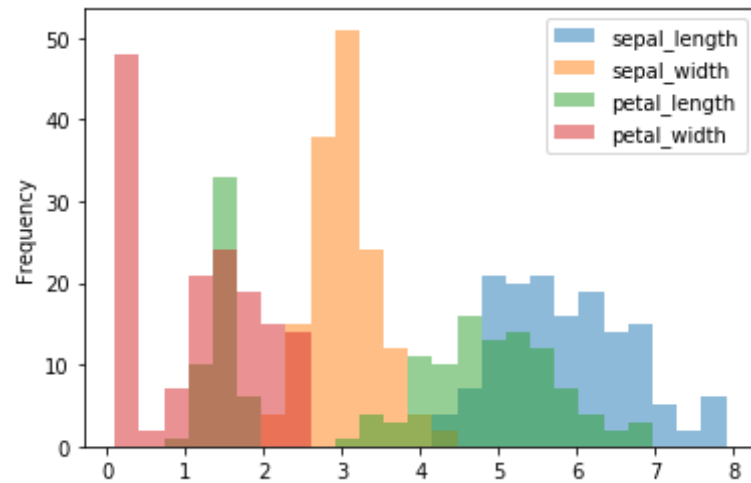


```
In [57]: fig,ax=plt.subplots()
         ax.hist(data.petal_length,bins=25)
```

```
Out[57]: (array([ 4., 19., 25.,  2.,  0.,  0.,  0.,  0.,  1.,  2.,  2.,  3.,  8.,
                9., 12.,  8., 13., 10.,  4., 12.,  5.,  5.,  2.,  1.,  3.]),
         array([1.    , 1.236, 1.472, 1.708, 1.944, 2.18 , 2.416, 2.652, 2.888,
                3.124, 3.36 , 3.596, 3.832, 4.068, 4.304, 4.54 , 4.776, 5.012,
                5.248, 5.484, 5.72 , 5.956, 6.192, 6.428, 6.664, 6.9  ]),
         <a list of 25 Patch objects>)
```

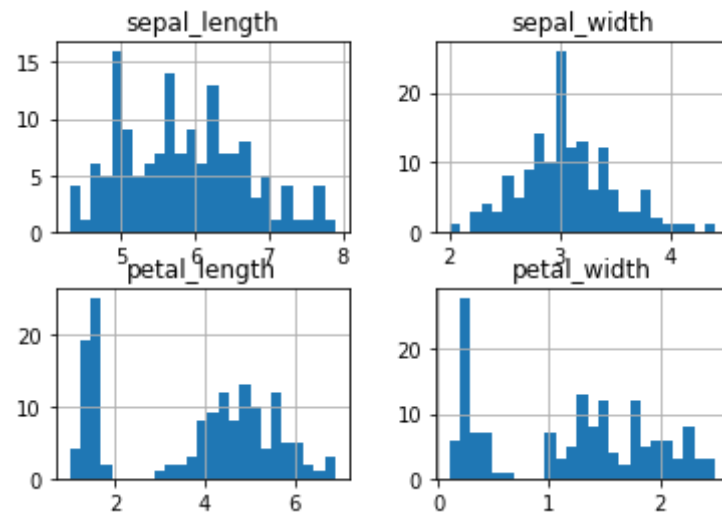


```
In [58]: ax=data.plot.hist(bins=25,alpha=0.5)
```



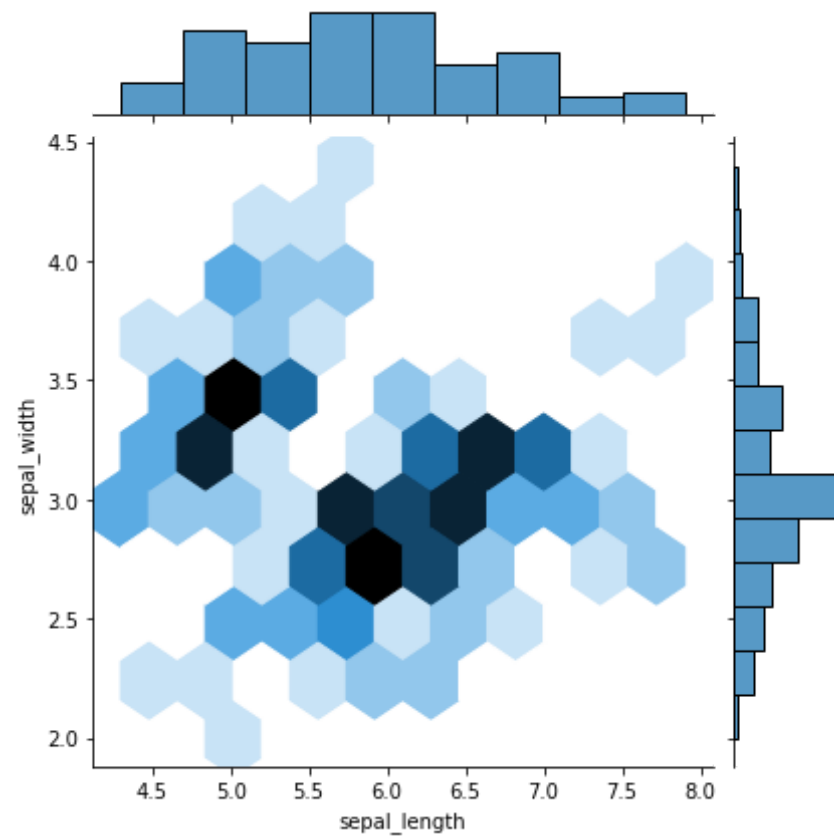
```
In [60]: data.hist(bins=25)
```

```
Out[60]: array([[<matplotlib.axes._subplots.AxesSubplot object at 0x7f27b542f410>,  
                <matplotlib.axes._subplots.AxesSubplot object at 0x7f27b57595d0>],  
               [<matplotlib.axes._subplots.AxesSubplot object at 0x7f27b583fc50>,  
                <matplotlib.axes._subplots.AxesSubplot object at 0x7f27b57efd90>]],  
              dtype=object)
```



```
In [61]: sns.jointplot(x=data['sepal_length'],y=data['sepal_width'],kind='hex')
```

```
Out[61]: <seaborn.axisgrid.JointGrid at 0x7f27b511eed0>
```



In []: