

DATA TOOLS

20
22

DR \ AMIRA EL SAYED



Team Members

Ali Hassan Ali (20201498637) \ Mohamed Farag Abdel Maqsoud (20201498657)
Mahmoud Ayman Mohamed (20201474689) \ Rewan Mohamed Abdo (20201373309)

- First, we will talk about our data set :
The iris data are a data frame of 150 measurements of iris petal and sepal lengths and widths, with 50 measurements for each species of “setosa,” “versicolor,” and “virginica.” Let us assume that we are doing some computation on the sepal length.



○ Second, deference between sepal and petal :

The main difference between sepals and petals is that sepals are the green color, leaf-like structures that form the outermost whorl whereas petals are the bright-colored petaloid structures which form the inner whorl. Furthermore, sepals are responsible for protecting the inner parts of the budding flower while petals are responsible for attracting pollinators and protecting reproductive structures

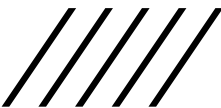




Third, we will talk about our columns:

SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm
In this column we have Length of 150 sepal's flower	In this column we have width of 150 sepal's flower	In this column we have length of 150 petal's flower	In this column we have width of 150 petal's flower

Let's Start with Python:




```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
import plotly.express as px

In [1]:
iris = sns.load_dataset("iris")
df=pd.DataFrame(iris)
data=df.copy()
df

Out[1]:
   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 [5]:
#visualization

In [3]:
#histogram for Sepal Length
plt.figure(figsize = (10, 7))
x = df.sepal_length
plt.hist(x, bins = 20, color = "blue")
plt.title("Sepal Length in cm")
plt.xlabel("Sepal_Length_cm")
plt.ylabel("Count")
plt.show()

Sepal_Length in cm

Count

In [7]:
#histogram for Sepal Width
plt.figure(figsize = (10, 7))
x = df.sepal_width
plt.hist(x, bins = 20, color = "blue")
plt.title("Sepal Width in cm")
plt.xlabel("Sepal_Width_cm")
plt.ylabel("Count")
plt.show()

Sepal Width in cm

Count

In [4]:
#histogram for Petal Length
plt.figure(figsize = (10, 7))
x = df.petal_length
plt.hist(x, bins = 20, color = "blue")
plt.title("Petal Length Cm")
plt.xlabel("Petal_Length Cm")
plt.ylabel("Count")
plt.show()

Petal_Length Cm

Count

In [5]:
#histogram for Petal Width
plt.figure(figsize = (10, 7))
x = df.petal_width
plt.hist(x, bins = 20, color = "blue")
plt.title("Petal Width Cm")
plt.xlabel("Petal_Width Cm")
plt.ylabel("Count")
plt.show()

Petal_Width Cm

Count

In [10]:
sns.jointplot(x="sepal_length", y="sepal_width", data=iris, hue="species", fit_reg=False, legend=False)

sepal_width Cm
sepal_length Cm
species
setosa
versicolor
virginica

In [11]:
sns.jointplot(x="petal_length", y="petal_width", data=iris, hue="species", fit_reg=False, legend=False)

petal_width Cm
petal_length Cm
species
setosa
versicolor
virginica

In [12]:
sns.jointplot(x="petal_length", y="sepal_length", data=iris, hue="species", fit_reg=False, legend=False)

sepal_length Cm
petal_length Cm
species
setosa
versicolor
virginica

In [13]:
sns.jointplot(x="petal_width", y="sepal_width", data=iris, hue="species", fit_reg=False, legend=False)

sepal_width Cm
petal_width Cm
species
setosa
versicolor
virginica

In [14]:
sns.jointplot(x="petal_length", y="sepal_width", data=iris, hue="species", fit_reg=False, legend=False)

sepal_width Cm
petal_length Cm
species
setosa
versicolor
virginica

In [15]:
sns.jointplot(x="sepal_length", y="petal_width", data=iris, hue="species", fit_reg=False, legend=False)

petal_width Cm
sepal_length Cm
species
setosa
versicolor
virginica

In [16]:
sepal_length=df.iloc[:,0].values
sepal_width=df.iloc[:,1].values
petal_length=df.iloc[:,2].values
petal_width=df.iloc[:,3].values
species=df.iloc[:,4].values
df = px.data.iris()
fig=px.scatter_3d(df, sepal_length,sepalwidth,petallength,
                  color="species")
fig.show()

In [17]:
df = px.data.iris()
fig=px.scatter_3d(df, sepal_length,sepalwidth, petalwidth,
                  color="species")
fig.show()

In [18]:
df = px.data.iris()
fig=px.scatter_3d(df, petallength,petalwidth, sepallength,
                  color="species")
fig.show()

In [19]:
df = px.data.iris()
fig=px.scatter_3d(df, petallength,petalwidth, sepalwidth,
                  color="species")
fig.show()

In [20]:
sns.jointplot(data=df, y=sepal_length, y=sepalwidth, hue=species, ratio=1, marginal_ticks=True)

sepal_width
sepal_length
species
setosa
versicolor
virginica

In [21]:
plt.figure(figsize = (10, 7))
sns.boxplot(x="species", y="sepal_length", data=df)

sepal_length
species
setosa
versicolor
virginica

In [22]:
plt.figure(figsize = (10, 7))
sns.boxplot(x="species", y="sepal_width", data=df)

sepal_width
species
setosa
versicolor
virginica

In [23]:
plt.figure(figsize = (10, 7))
sns.boxplot(x="species", y="petal_length", data=df)

petal_length
species
setosa
versicolor
virginica

In [24]:
plt.figure(figsize = (10, 7))
sns.boxplot(x="species", y="petal_width", data=df)

petal_width
species
setosa
versicolor
virginica

In [22]:
corr=df.corr()
fig, ax = plt.subplots(figsize=(5,4))
sns.heatmap(corr, annot=True, ax=ax, cmap = "coolwarm")
plt.show()

sepal_length  sepal_width  petal_length  petal_width  species_id
sepal_length  1.000000  0.81156  0.81629  0.81629  0.76909
sepal_width   0.81156  1.00000  0.43301  0.36633  0.42867
petal_length  0.81629  0.43301  1.00000  0.96336  0.95951
petal_width   0.81629  0.36633  0.96336  1.00000  0.95951
species_id    0.76909  0.42867  0.95951  0.95951  1.00000
```

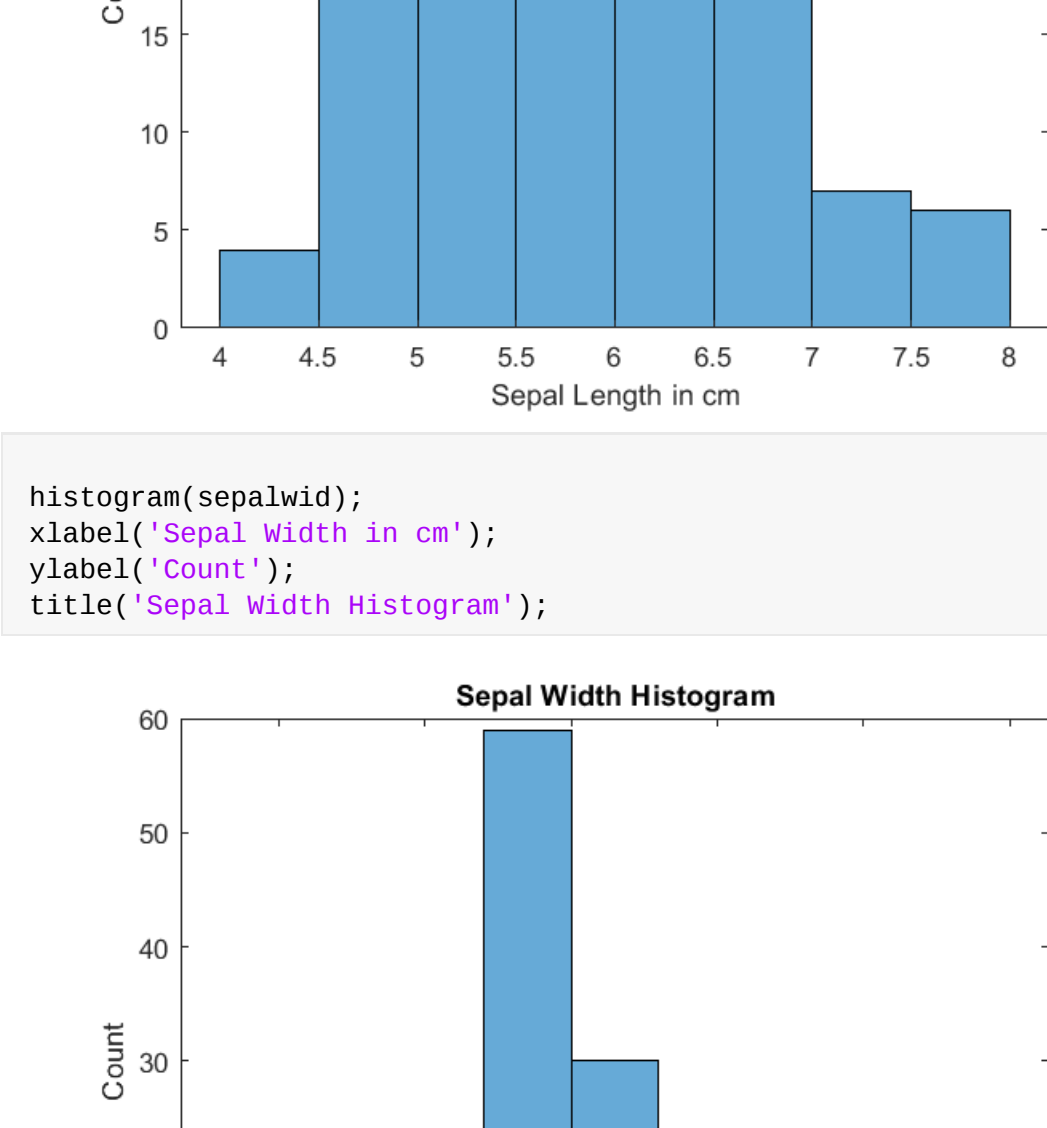
Then Matlab:


```
load fisheriris
ds = mat2dataset(meas);
ds(1:10,:)
```

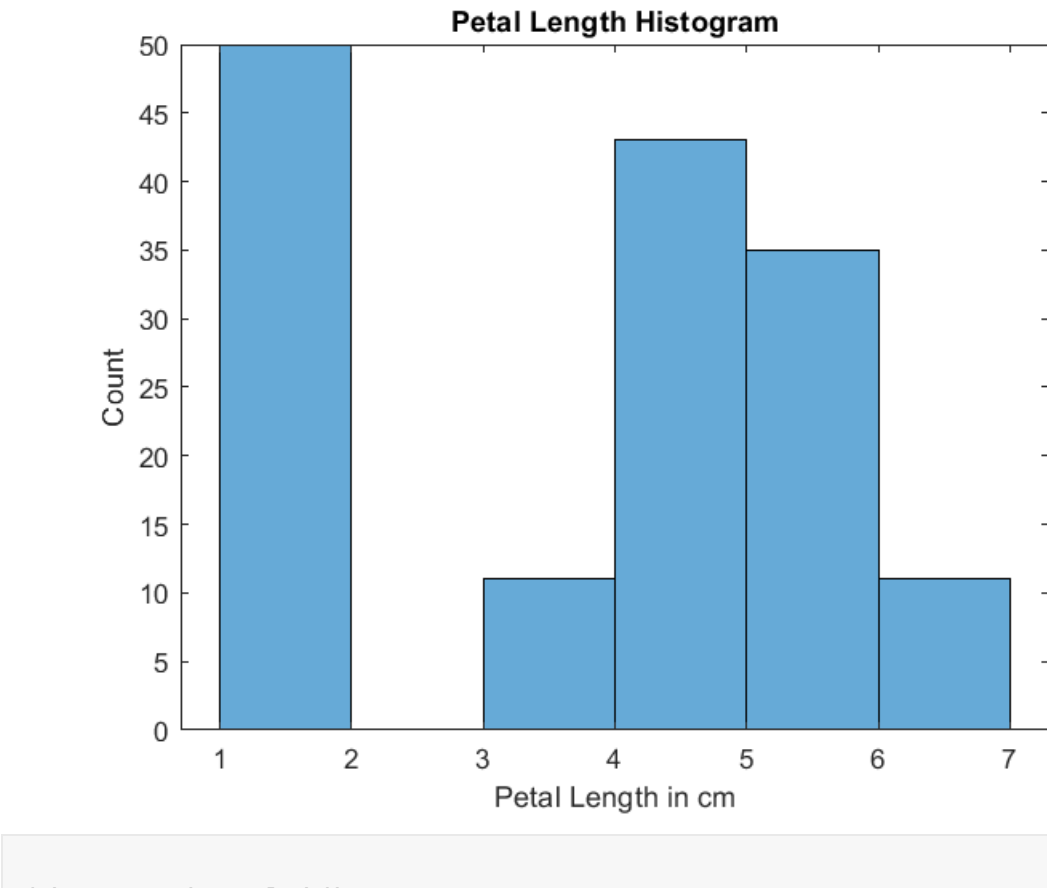
```
ans =
    meas1    meas2    meas3    meas4
    5.1      3.5      1.4      0.2
    4.9      3      1.4      0.2
    4.7      3.2      1.3      0.2
    4.6      3.1      1.5      0.2
    5      3.6      1.4      0.2
    5.4      3.9      1.7      0.4
    4.8      3.4      1.4      0.3
    5      3.4      1.5      0.2
    4.4      2.9      1.4      0.2
    4.9      3.1      1.5      0.1
```

```
sepalen = meas(:,1);
sepalwid = meas(:,2);
petallen = meas(:,3);
petalwid = meas(:,4);

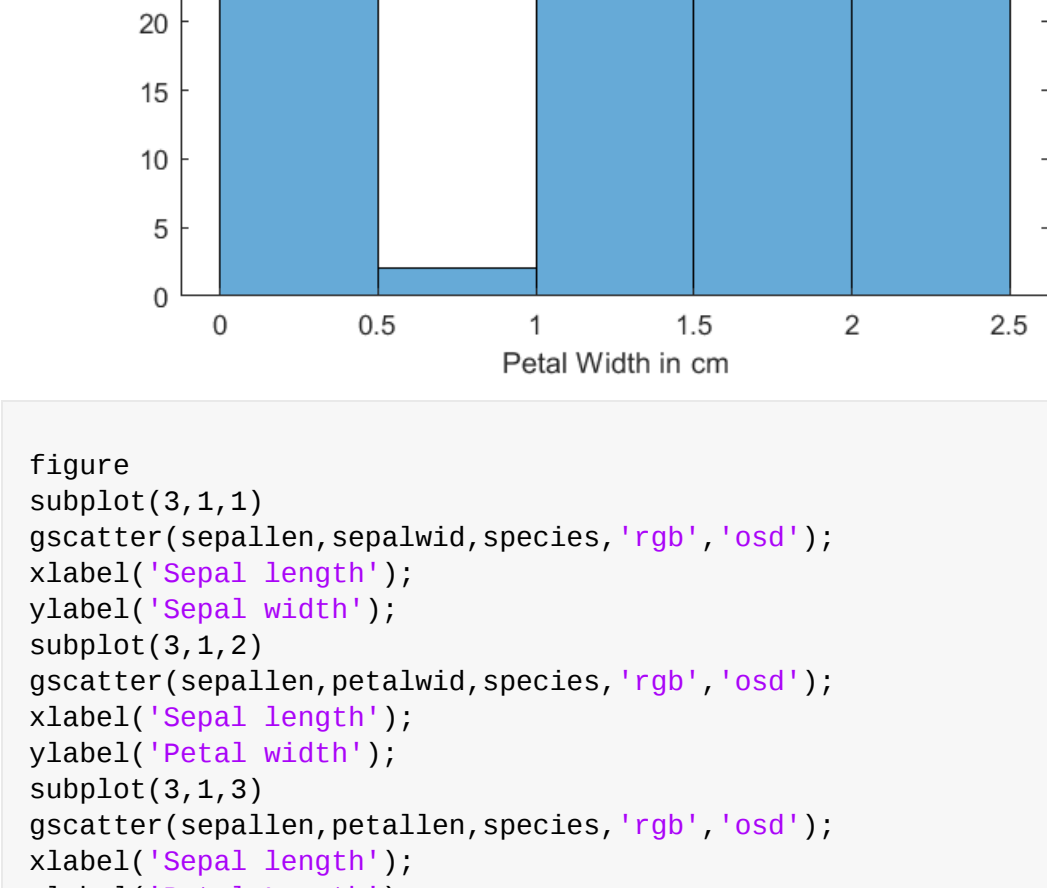
figure
histogram(sepalen);
xlabel('Sepal Length in cm');
ylabel('Count');
title('Sepal Length Histogram');
```



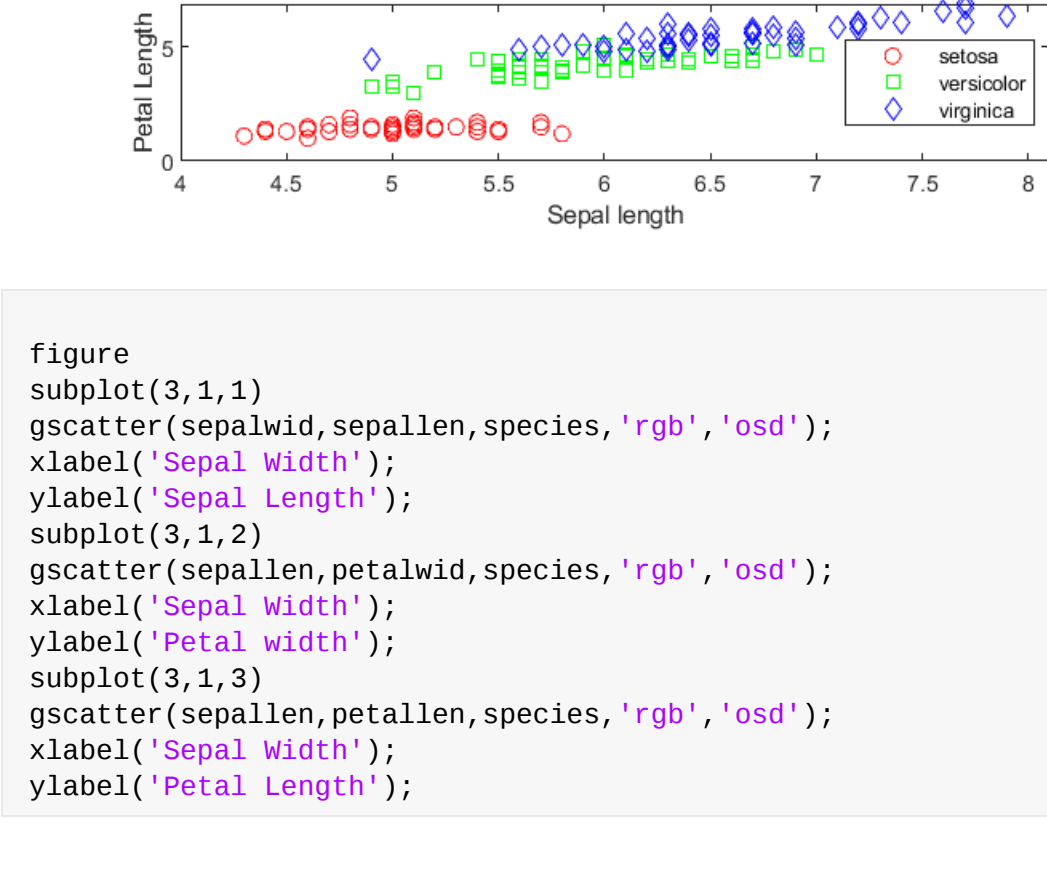
```
histogram(sepalwid);
xlabel('Sepal Width in cm');
ylabel('Count');
title('Sepal Width Histogram');
```



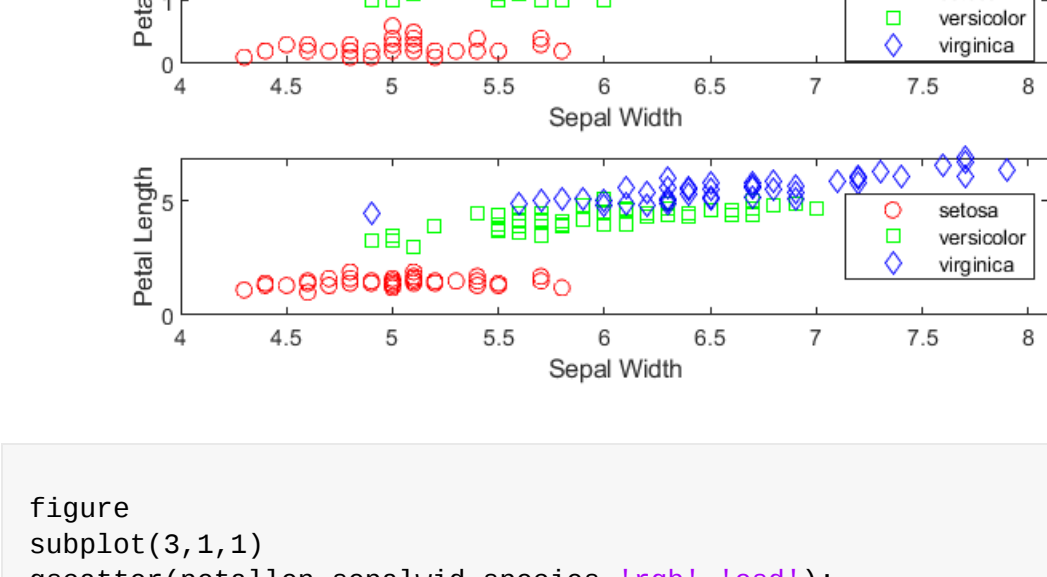
```
histogram(petalen);
xlabel('Petal Length in cm');
ylabel('Count');
title('Petal Length Histogram');
```



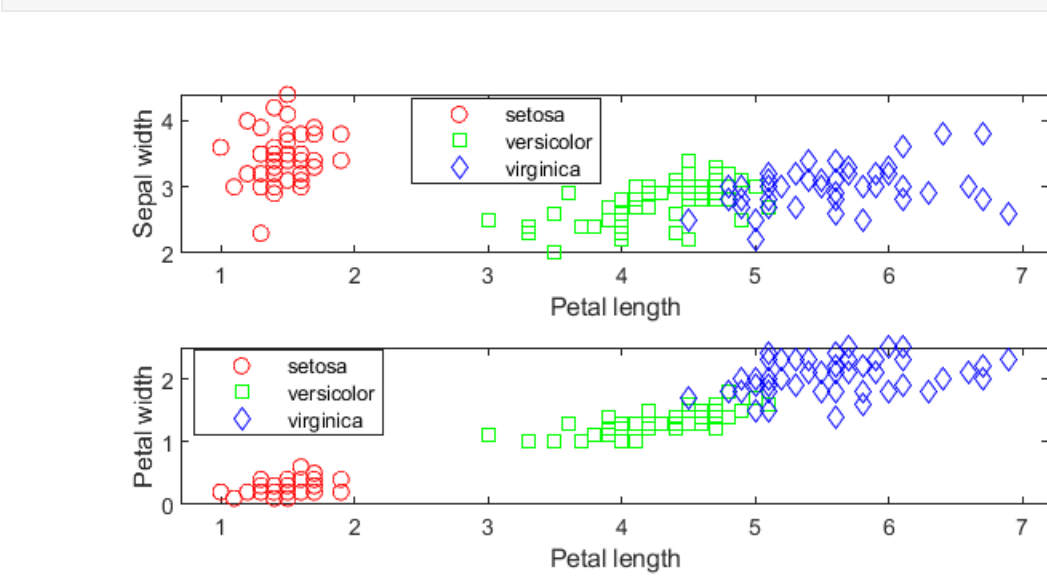
```
histogram(petalwid);
xlabel('Petal Width in cm');
ylabel('Count');
title('Petal Width Histogram');
```



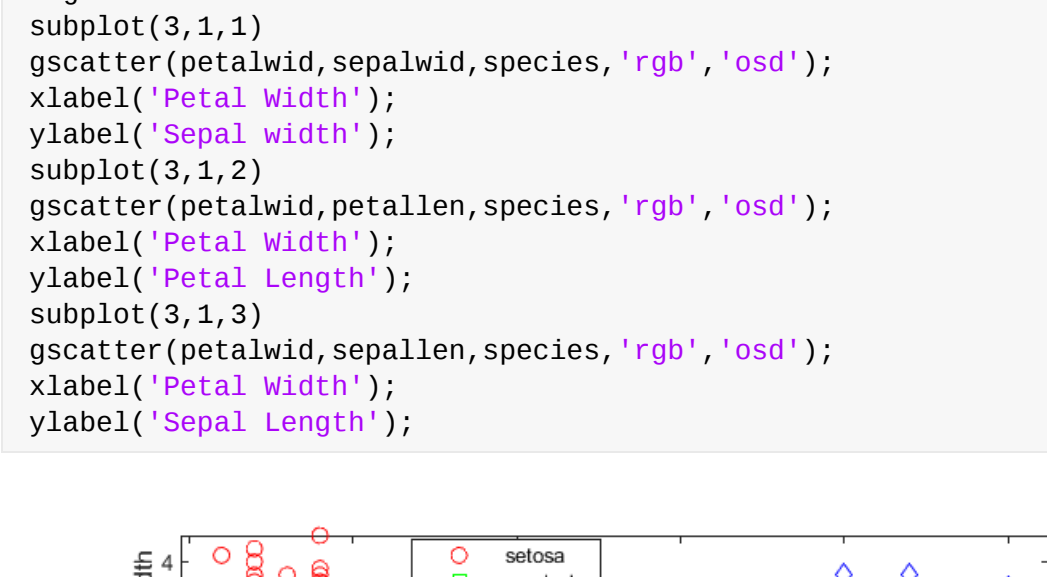
```
figure
subplot(3,1,1)
gscatter(sepalen,sepalwid,species,'rgb','osd');
xlabel('Sepal Length');
ylabel('Sepal Width');
subplot(3,1,2)
gscatter(sepalen,petalwid,species,'rgb','osd');
xlabel('Sepal Length');
ylabel('Petal Width');
subplot(3,1,3)
gscatter(sepalen,petallen,species,'rgb','osd');
xlabel('Sepal Length');
ylabel('Petal Length');
```



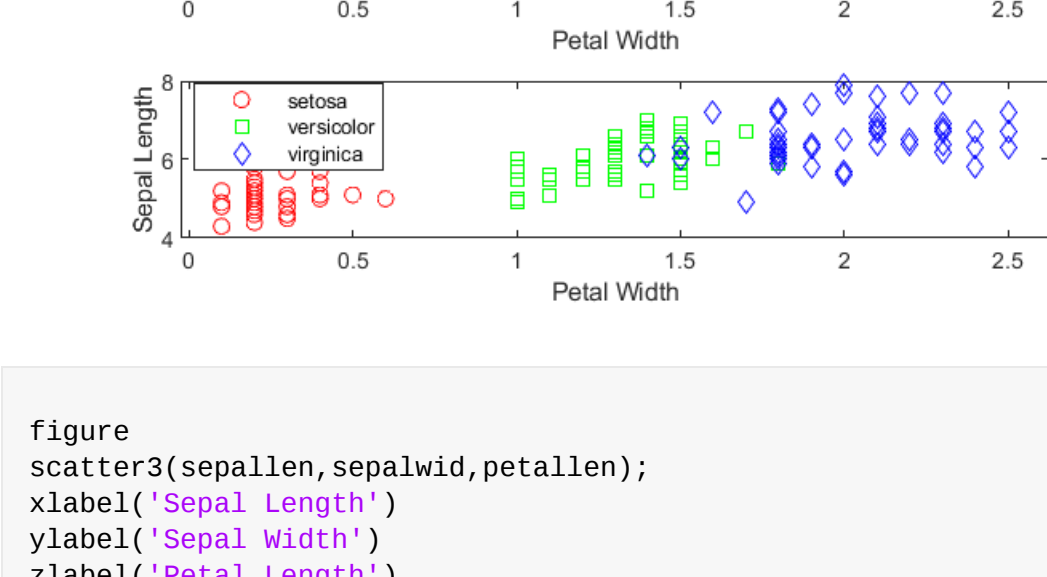
```
figure
subplot(3,1,1)
gscatter(sepalwid,sepalen,species,'rgb','osd');
xlabel('Sepal Width');
ylabel('Sepal Length');
subplot(3,1,2)
gscatter(sepalwid,petalwid,species,'rgb','osd');
xlabel('Sepal Width');
ylabel('Petal Width');
subplot(3,1,3)
gscatter(sepalwid,petallen,species,'rgb','osd');
xlabel('Sepal Width');
ylabel('Petal Length');
```



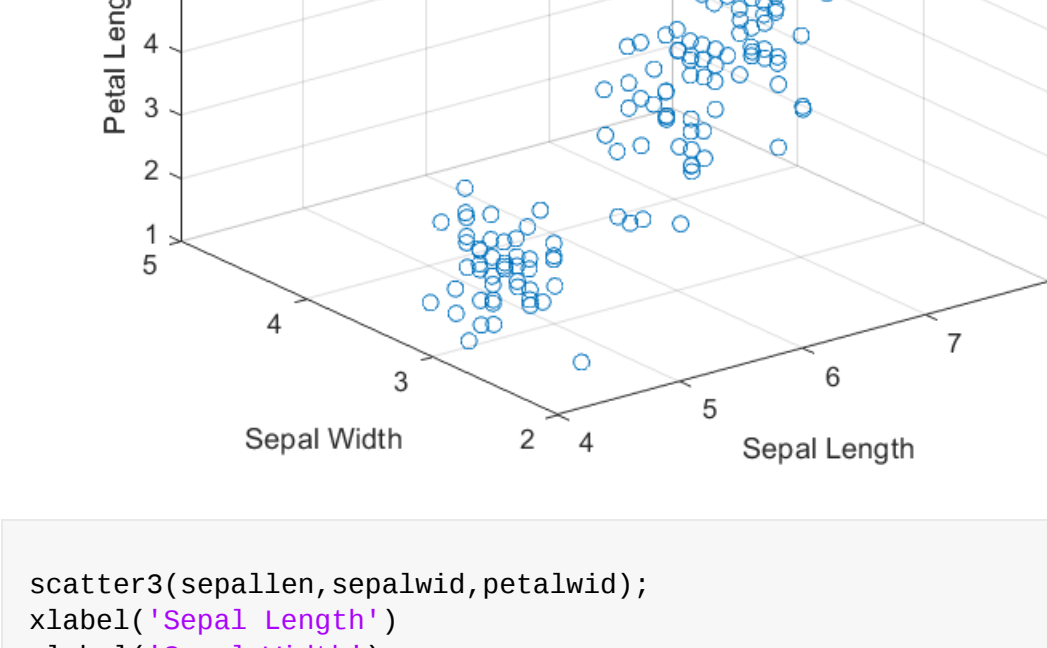
```
figure
subplot(3,1,1)
gscatter(petalen,sepalwid,species,'rgb','osd');
xlabel('Petal Length');
ylabel('Sepal Width');
subplot(3,1,2)
gscatter(petalen,petalwid,species,'rgb','osd');
xlabel('Petal Length');
ylabel('Petal Width');
subplot(3,1,3)
gscatter(petalen,petallen,species,'rgb','osd');
xlabel('Petal Length');
ylabel('Sepal Length');
```



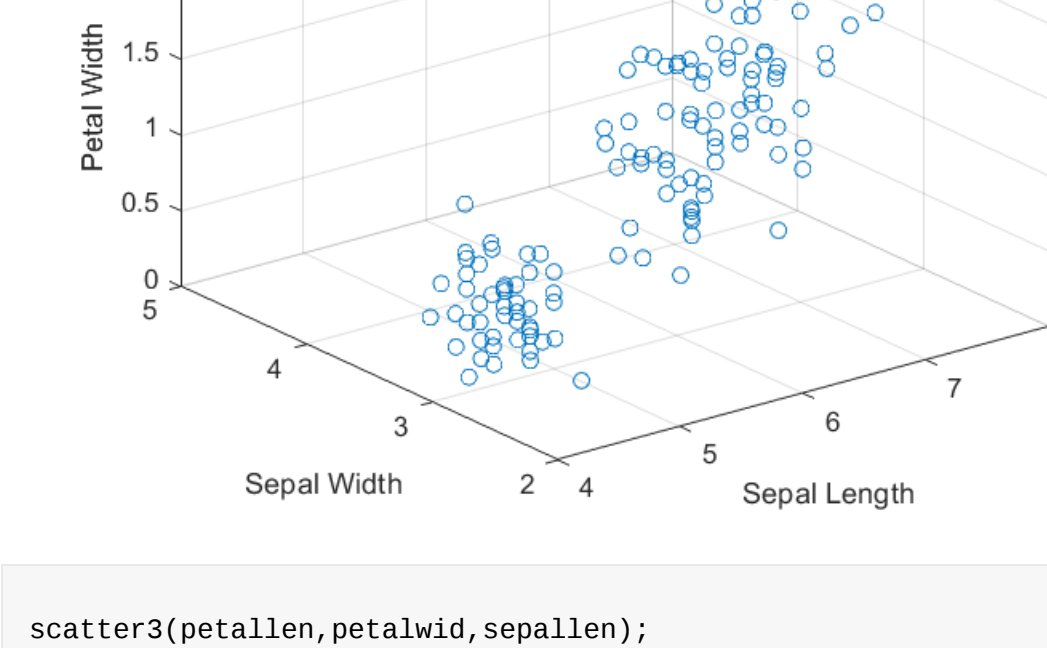
```
figure
subplot(3,1,1)
gscatter(petalwid,sepalwid,species,'rgb','osd');
xlabel('Petal Width');
ylabel('Sepal Width');
subplot(3,1,2)
gscatter(petalwid,petallen,species,'rgb','osd');
xlabel('Petal Width');
ylabel('Sepal Length');
```



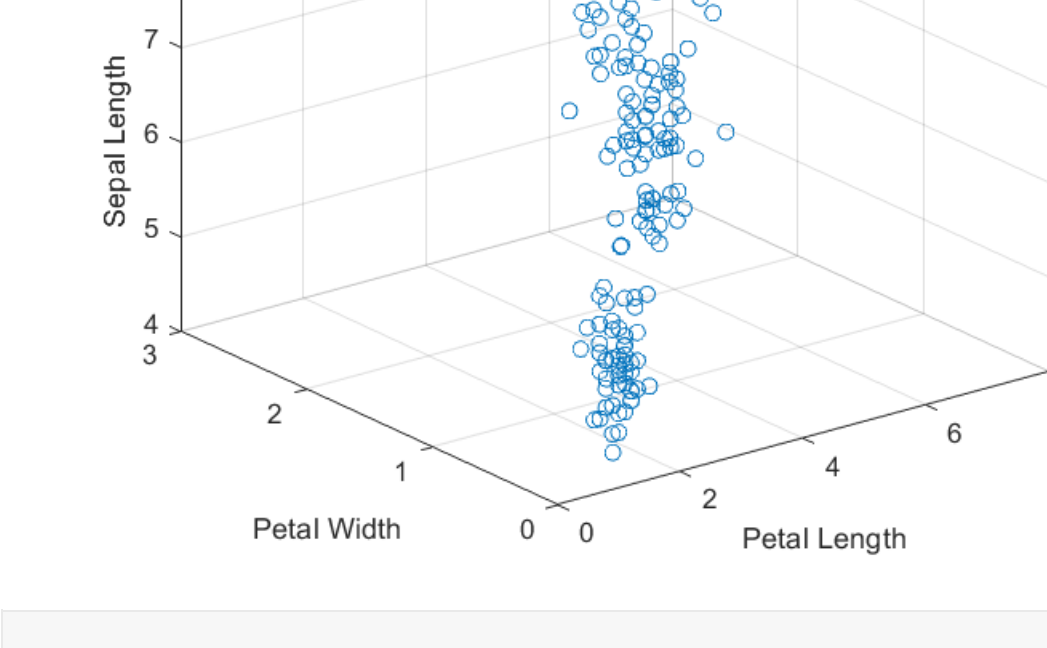
```
figure
scatter3(sepalen,sepalwid,petallen);
xlabel('Sepal Length');
ylabel('Sepal Width');
zlabel('Petal Length');
```



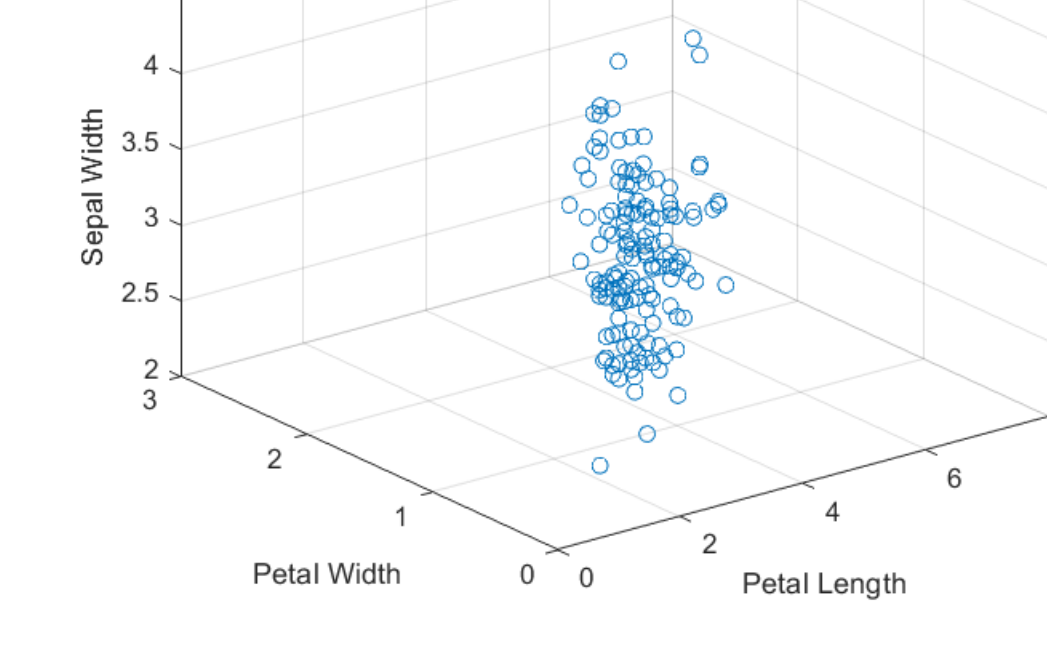
```
scatter3(sepalen,sepalwid,petalwid);
xlabel('Sepal Length');
ylabel('Sepal Width');
zlabel('Petal Width');
```



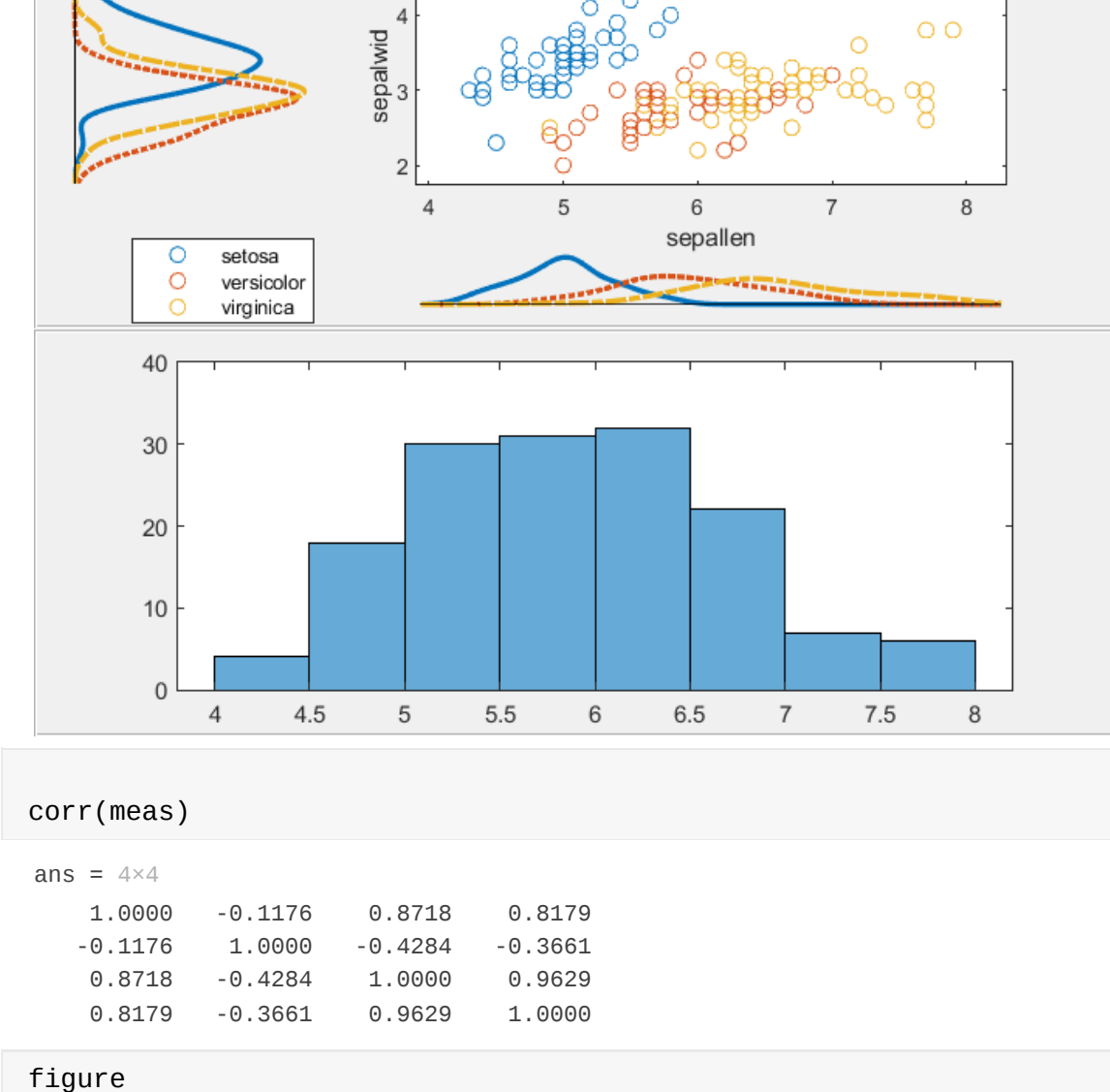
```
scatter3(petalen,petalwid,sepalen);
xlabel('Petal Length');
ylabel('Petal Width');
zlabel('Sepal Length');
```



```
scatter3(petalen,petalwid,sepalwid);
xlabel('Petal Length');
ylabel('Petal Width');
zlabel('Sepal Width');
```



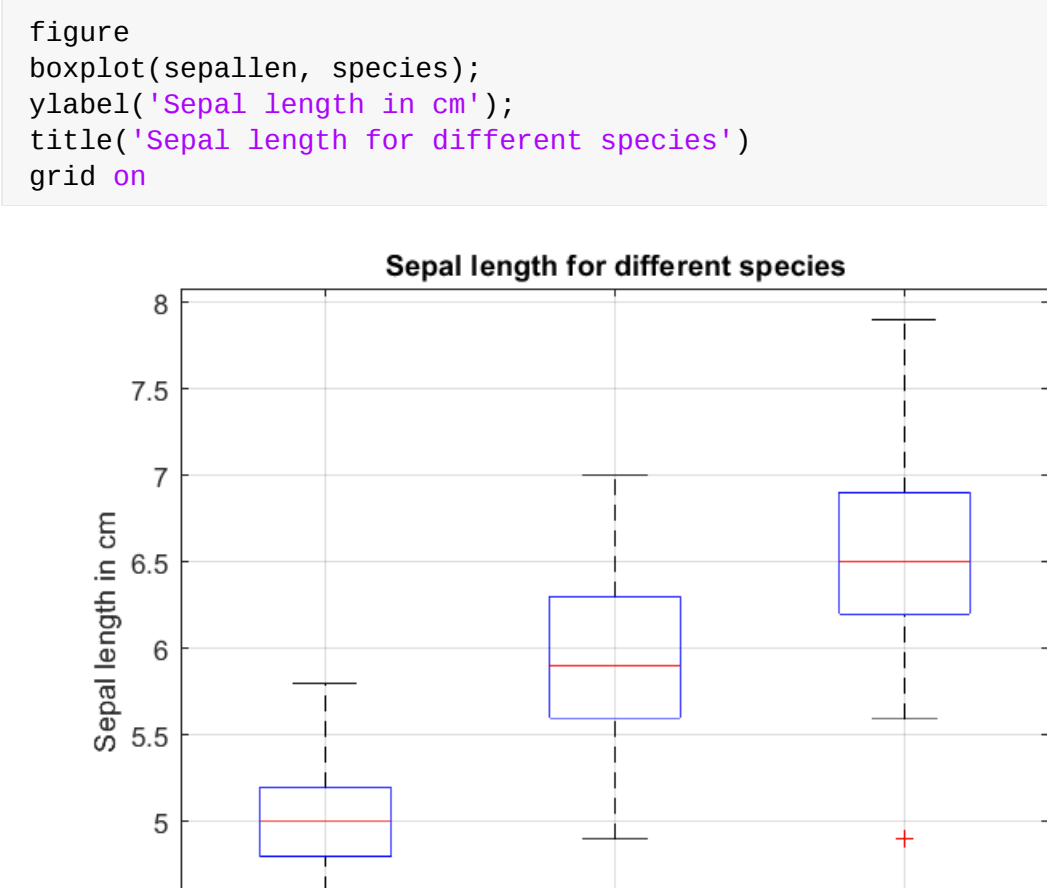
```
figure
hpl = uipanel('position',[0 5 1 5]);
hp2 = uipanel('position',[0 0 1 5]);
scatterhist(sepalen,sepalwid,'Group',species,'kernel','on','Parent',hpl);
axes('Parent',hp2);
histogram(sepalen);
```



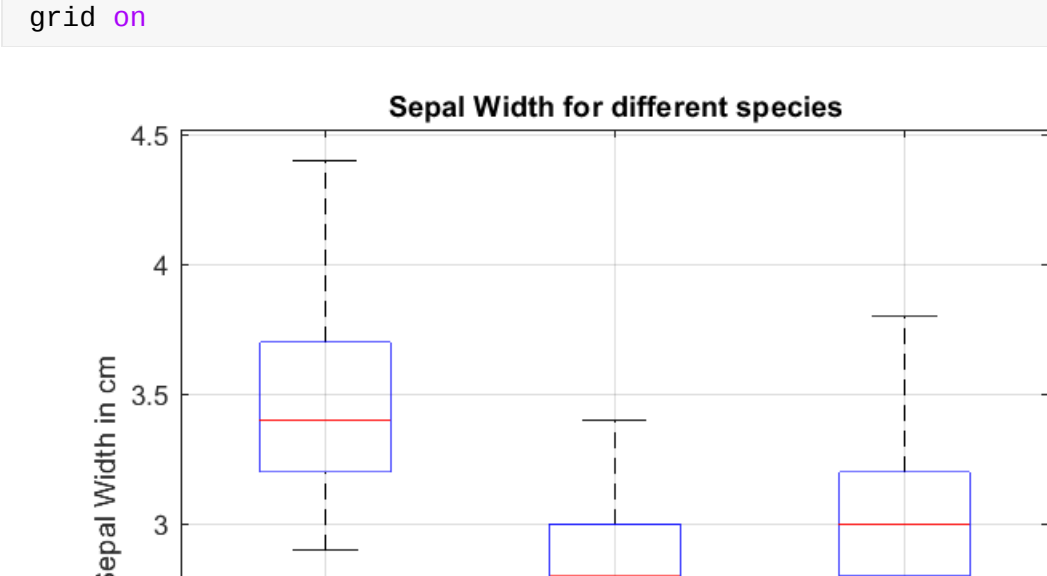
```
corr(meas)
```

```
ans =
    -0.4
    -0.1176    0.8718    0.8179
    1.0000    -0.4284    -0.3661
    -0.1176    1.0000    0.9629
    0.8718    -0.4284    1.0000
    0.8179    -0.3661    0.9629    1.0000
```

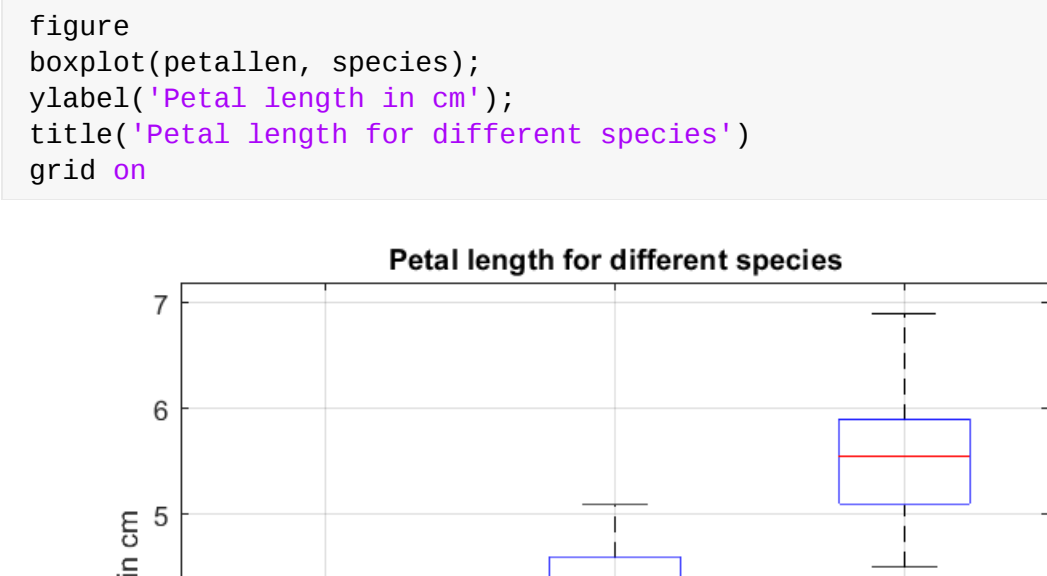
```
figure
labels = {'Sepal Length','Sepal Width','Petal Length','Petal Width'};
heatmap(labels,labels,corr(meas));
```



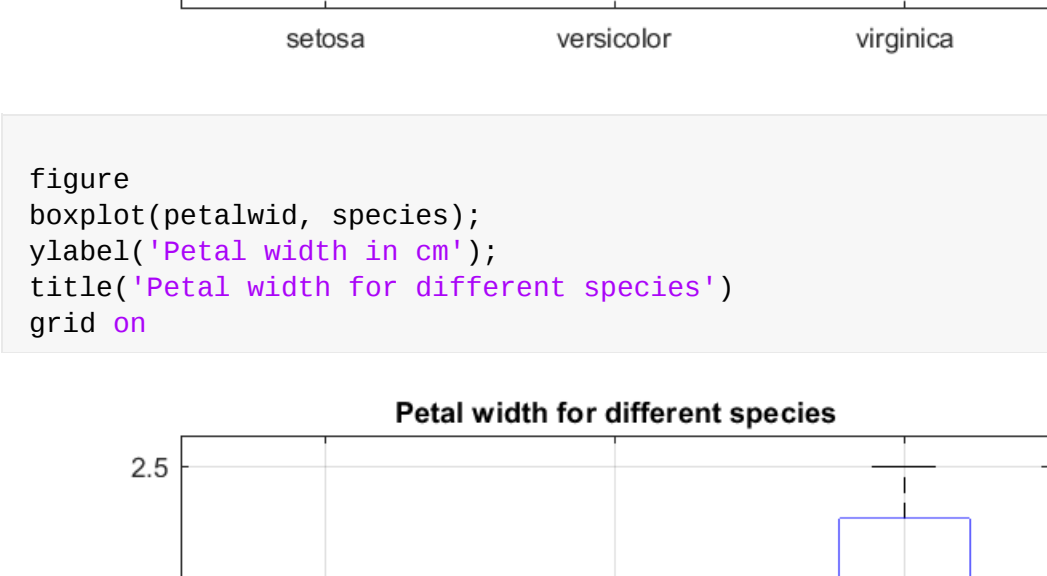
```
figure
boxplot(sepalen,species);
xlabel('Sepal Length in cm');
title('Sepal Length for different species')
grid on
```



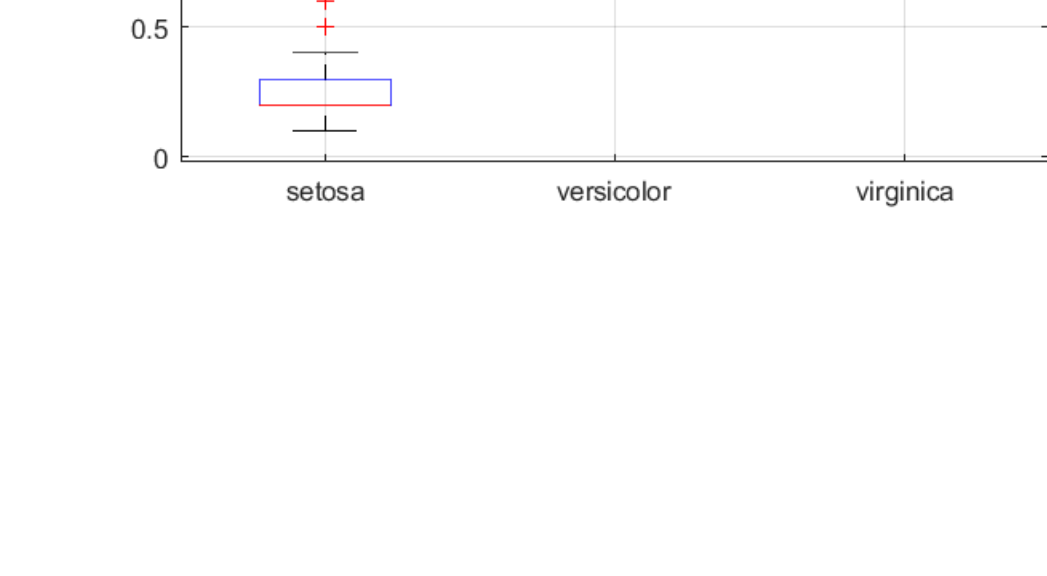
```
figure
boxplot(sepalwid,species);
xlabel('Sepal Width in cm');
title('Sepal Width for different species')
grid on
```



```
figure
boxplot(petalen,species);
xlabel('Petal Length in cm');
title('Petal Length for different species')
grid on
```



```
figure
boxplot(petalwid,species);
xlabel('Petal width in cm');
title('Petal width for different species')
grid on
```



○ Conclusion :

- After this amazing journey in data

We prefer python because :

- 1.It open source language
- 2.it's totally free
- 3.Easy to tracing code
- 4.Easy to debugging code
- 5.Easy to learn
- 6.We can use it in many domains

