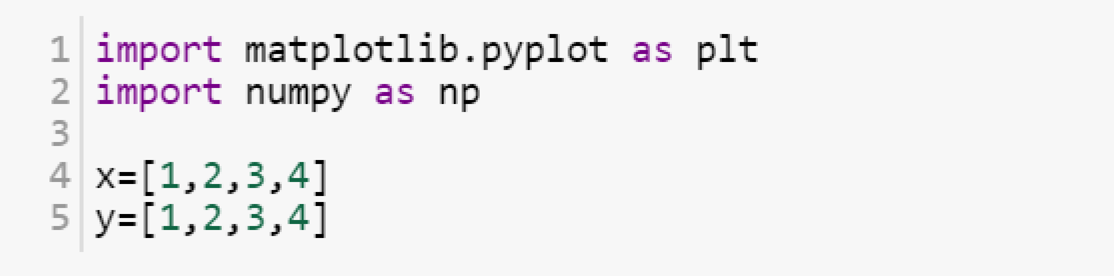
Data Visualization

1. Matplotlib
2. Seaborn

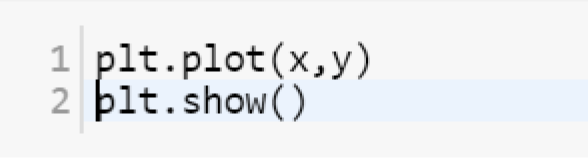
Matplotlib

- Line Chart Plotting :

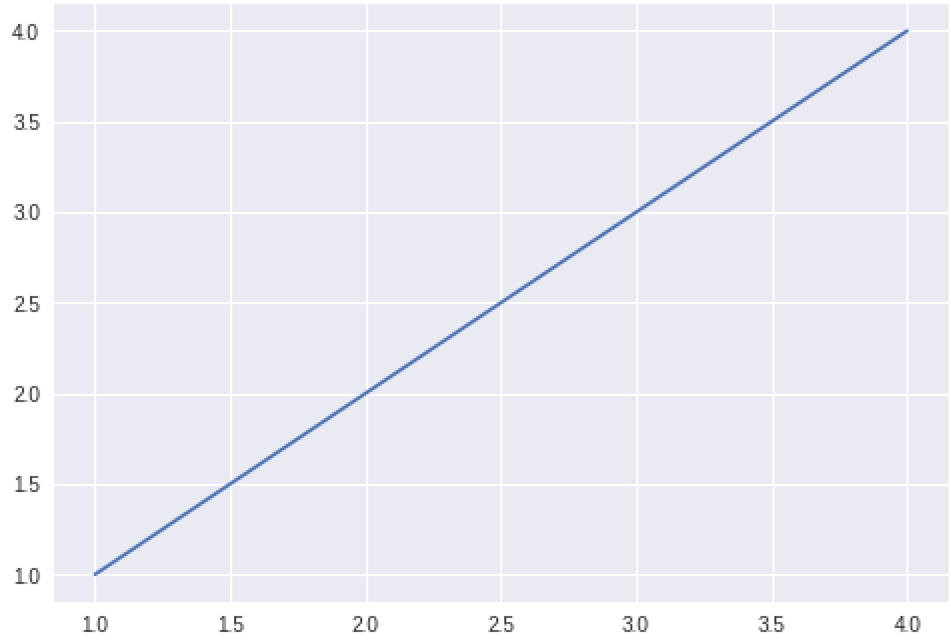


```
import matplotlib.pyplot as plt
import numpy as np

x=[1,2,3,4]
y=[1,2,3,4]
```

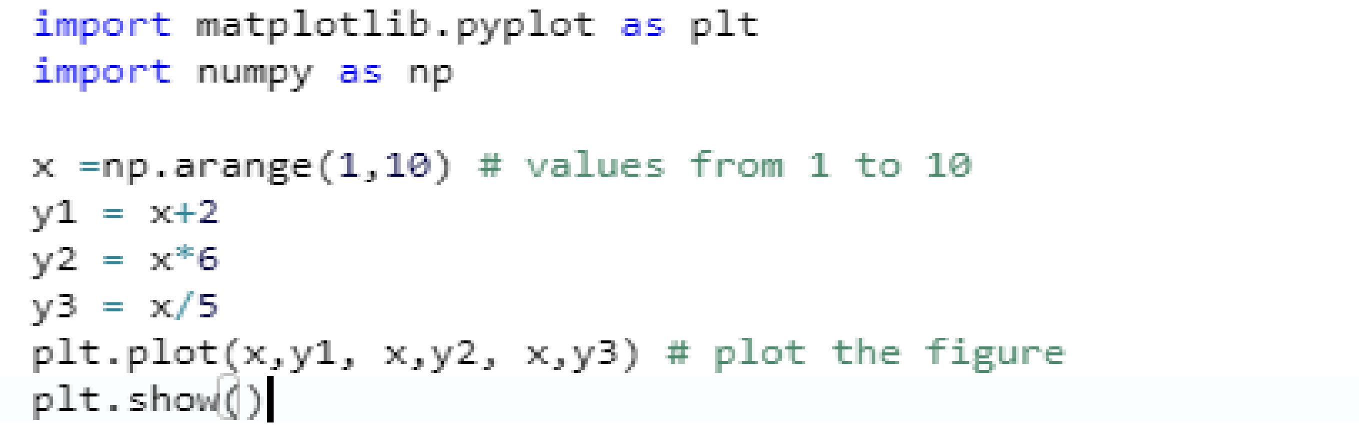


```
plt.plot(x,y)
plt.show()
```



Multiple Lines Plotting on the Same Graph

- Matplotlib has also the ability to plot multiple numbers of lines on the same graph.
- The method plot() method can contains many lines. Like plot(x,y1, x,y2,x,y3···).

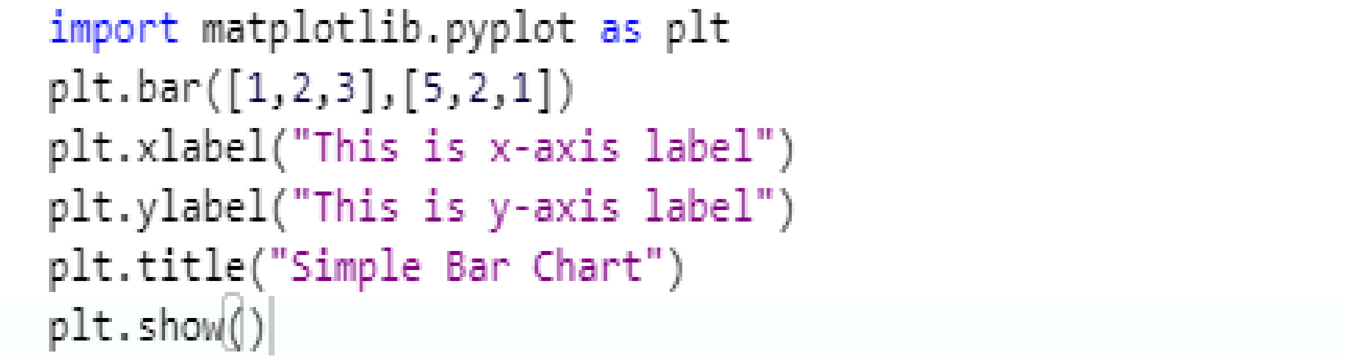


```
import matplotlib.pyplot as plt
import numpy as np

x =np.arange(1,10) # values from 1 to 10
y1 = x+2
y2 = x*6
y3 = x/5
plt.plot(x,y1, x,y2, x,y3) # plot the figure
plt.show()
```

Bar Chart Plotting

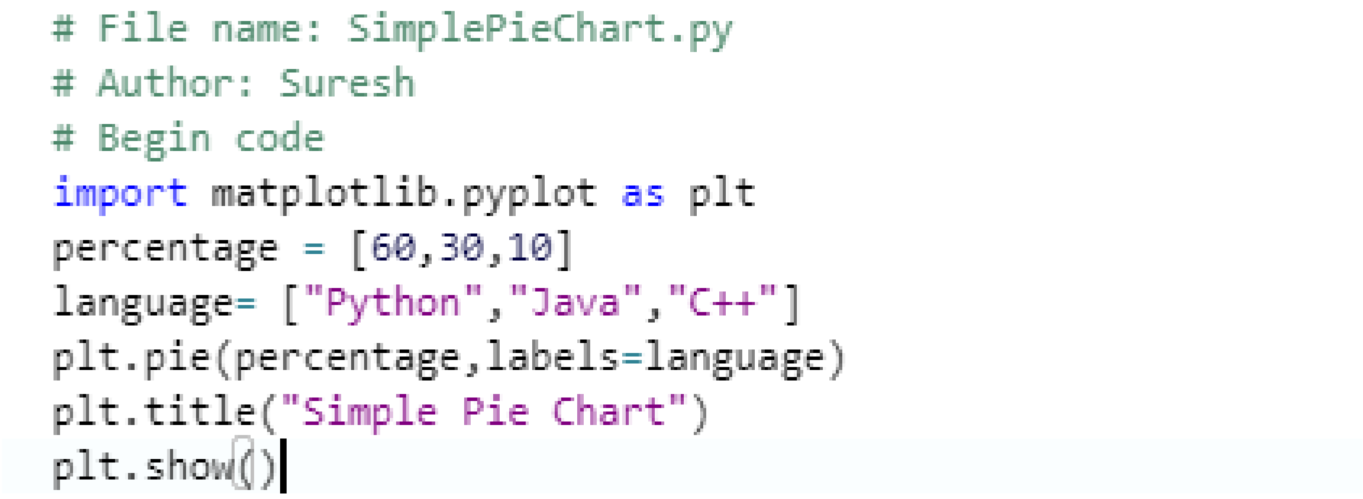
- You will use the bar() method for plotting the bar chart.
- The **plt.bar([x1,x2,x3···],[y1,y2,y3···])**
- x1,x2,x3··· are values on the x–axis and y1,y2,y.. are height of bar chart.



```
import matplotlib.pyplot as plt
plt.bar([1,2,3],[5,2,1])
plt.xlabel("This is x-axis label")
plt.ylabel("This is y-axis label")
plt.title("Simple Bar Chart")
plt.show()
```

Pie Chart Plotting

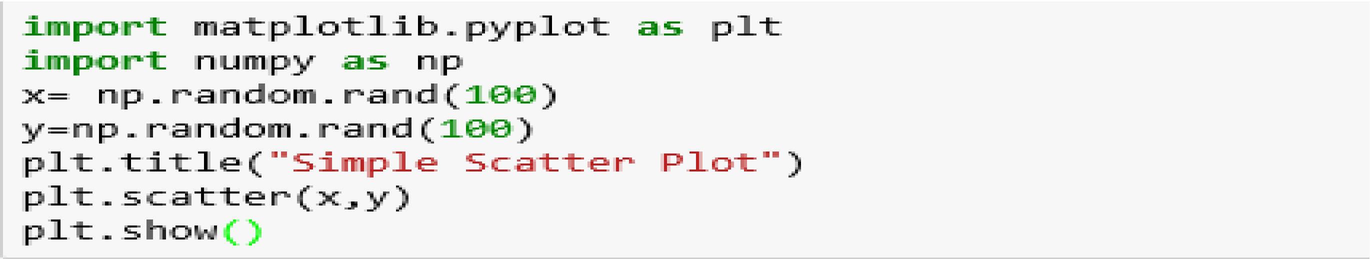
- The **pie()** function in the Maplotlib provides you to plot pie chart from an array x.
- **percentage = [60,30,10]**
- **language= [“Python”,”Java”,”C++”]**



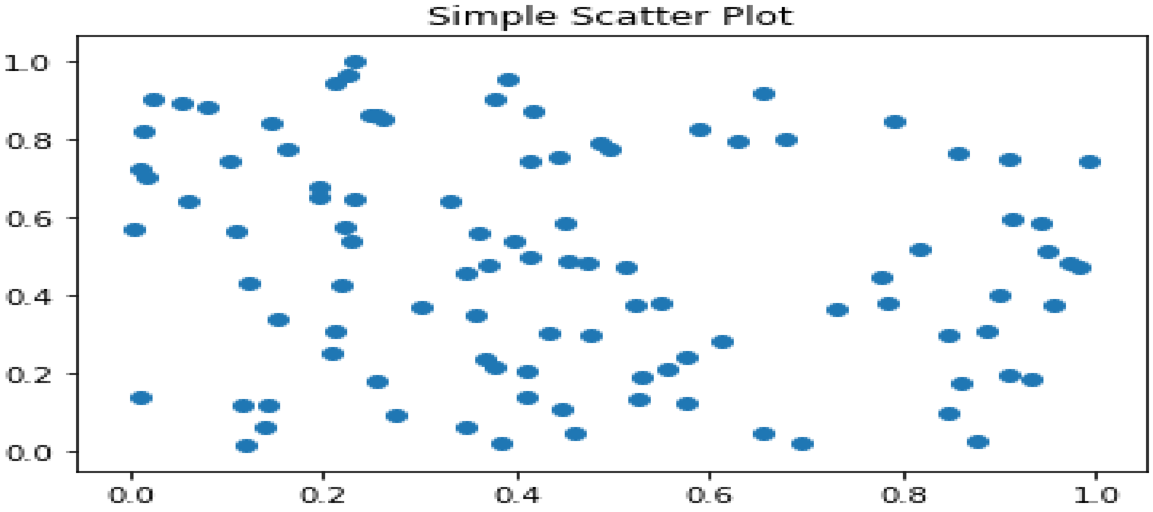
```
# File name: SimplePieChart.py
# Author: Suresh
# Begin code
import matplotlib.pyplot as plt
percentage = [60,30,10]
language= ["Python","Java","C++"]
plt.pie(percentage,labels=language)
plt.title("Simple Pie Chart")
plt.show()
```

Scatter plot

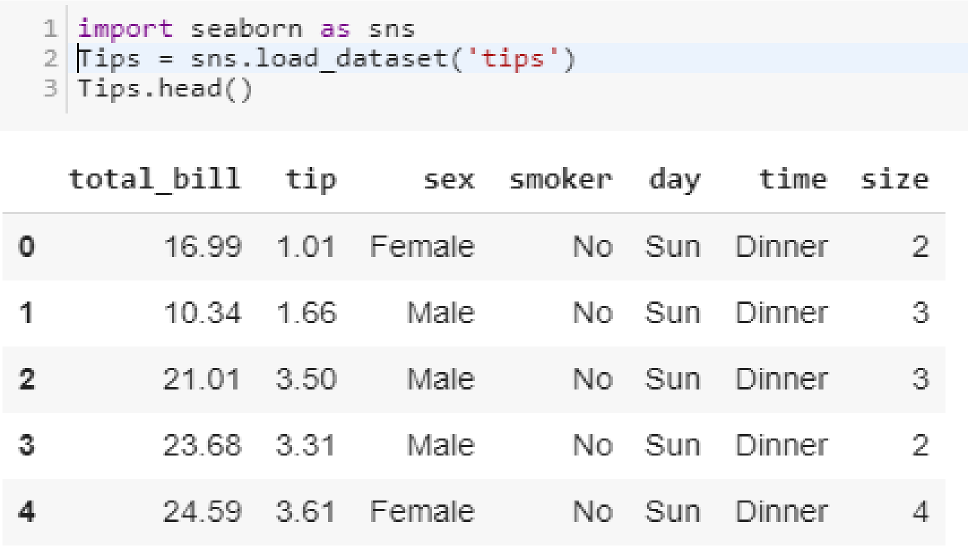
- A scatter plot just shows one point for each dataset entry.



```
import matplotlib.pyplot as plt
import numpy as np
x= np.random.rand(100)
y=np.random.rand(100)
plt.title("Simple Scatter Plot")
plt.scatter(x,y)
plt.show()
```

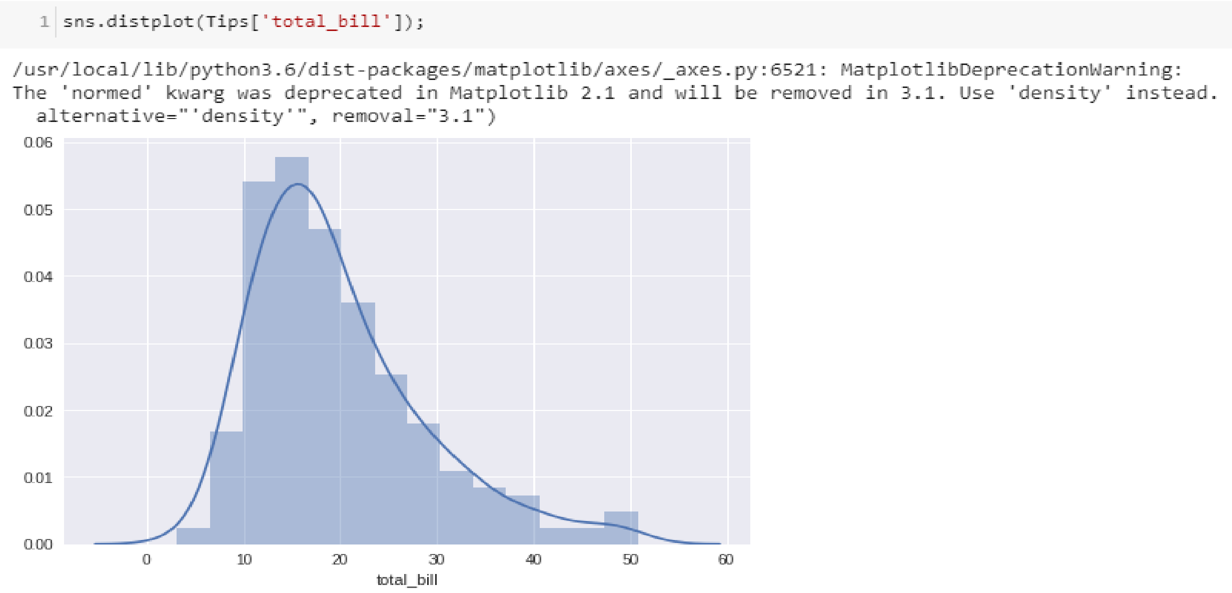


Seaborn



```python
import seaborn as sns
Tips = sns.load_dataset('tips')
Tips.head()
```

	total_bill	tip	sex	smoker	day	time	size
0	16.99	1.01	Female	No	Sun	Dinner	2
1	10.34	1.66	Male	No	Sun	Dinner	3
2	21.01	3.50	Male	No	Sun	Dinner	3
3	23.68	3.31	Male	No	Sun	Dinner	2
4	24.59	3.61	Female	No	Sun	Dinner	4

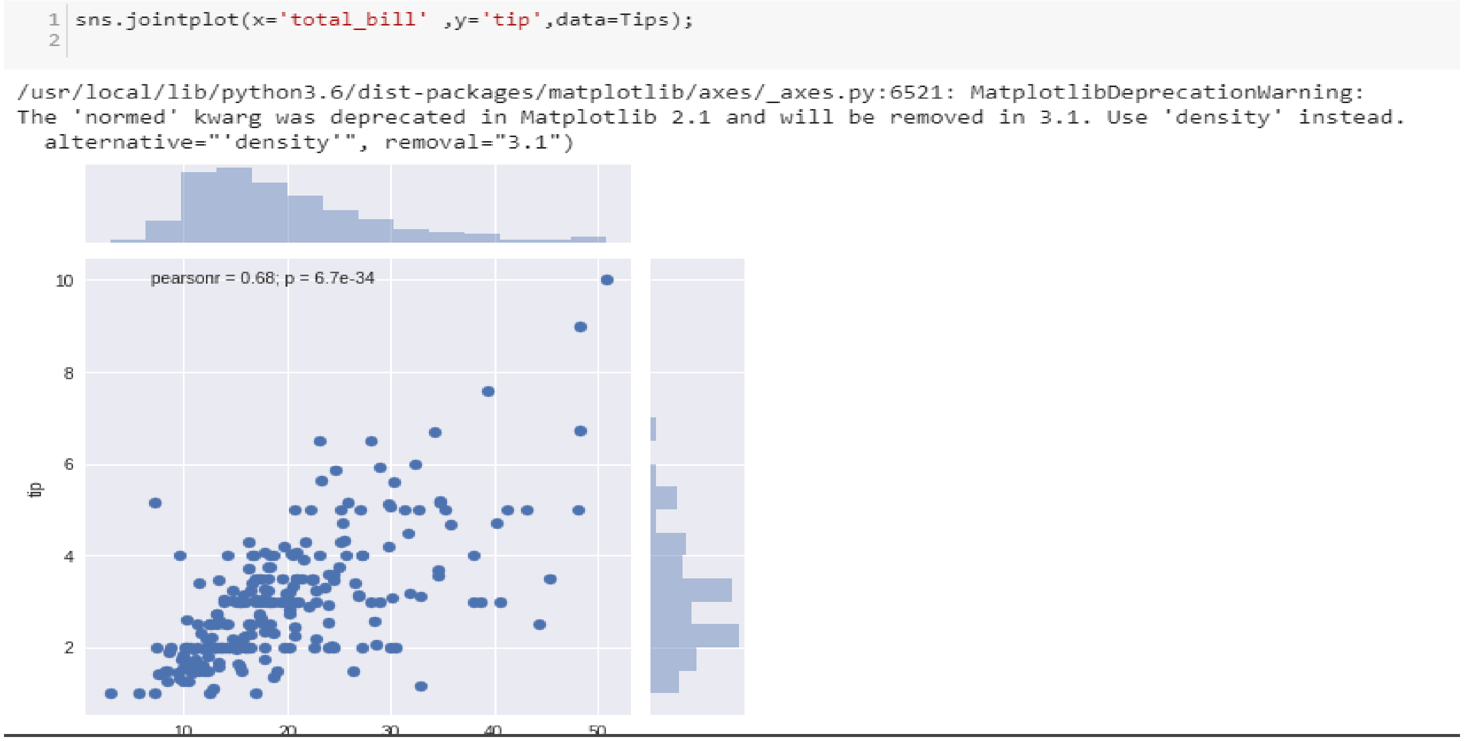


```python
sns.distplot(Tips['total_bill']);
```

```
/usr/local/lib/python3.6/dist-packages/matplotlib/axes/_axes.py:6521: MatplotlibDeprecationWarning:
The 'normed' kwarg was deprecated in Matplotlib 2.1 and will be removed in 3.1. Use 'density' instead.
  alternative="'density'", removal="3.1")
```

Joint Plot.

- Joint plot is bi variant., we can validate how the two variables are distributed .
- Sns.Jointplot(x= ‘total_bill’ ,y= ‘tip’ ,data=tips)



```
sns.jointplot(x='total_bill' ,y='tip',data=Tips);

```

```
/usr/local/lib/python3.6/dist-packages/matplotlib/axes/_axes.py:6521: MatplotlibDeprecationWarning:
The 'normed' kwarg was deprecated in Matplotlib 2.1 and will be removed in 3.1. Use 'density' instead.
  alternative="'density'", removal="3.1")
```

Pair plot

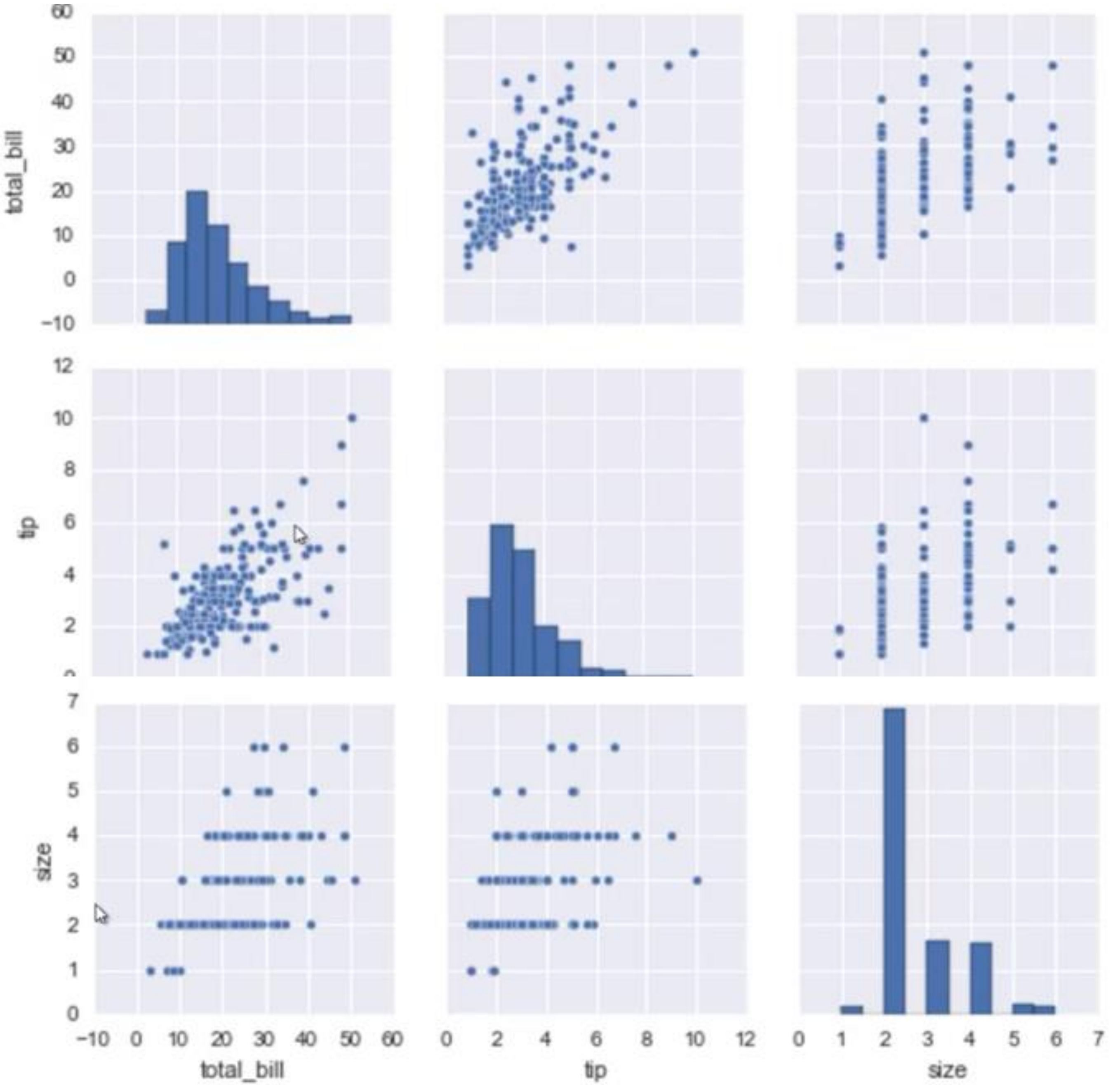
- Pair plot will plot a diagram all against the all possible combination of all numerical values. This is most import plots.

- We will work on this plot through out our ML.

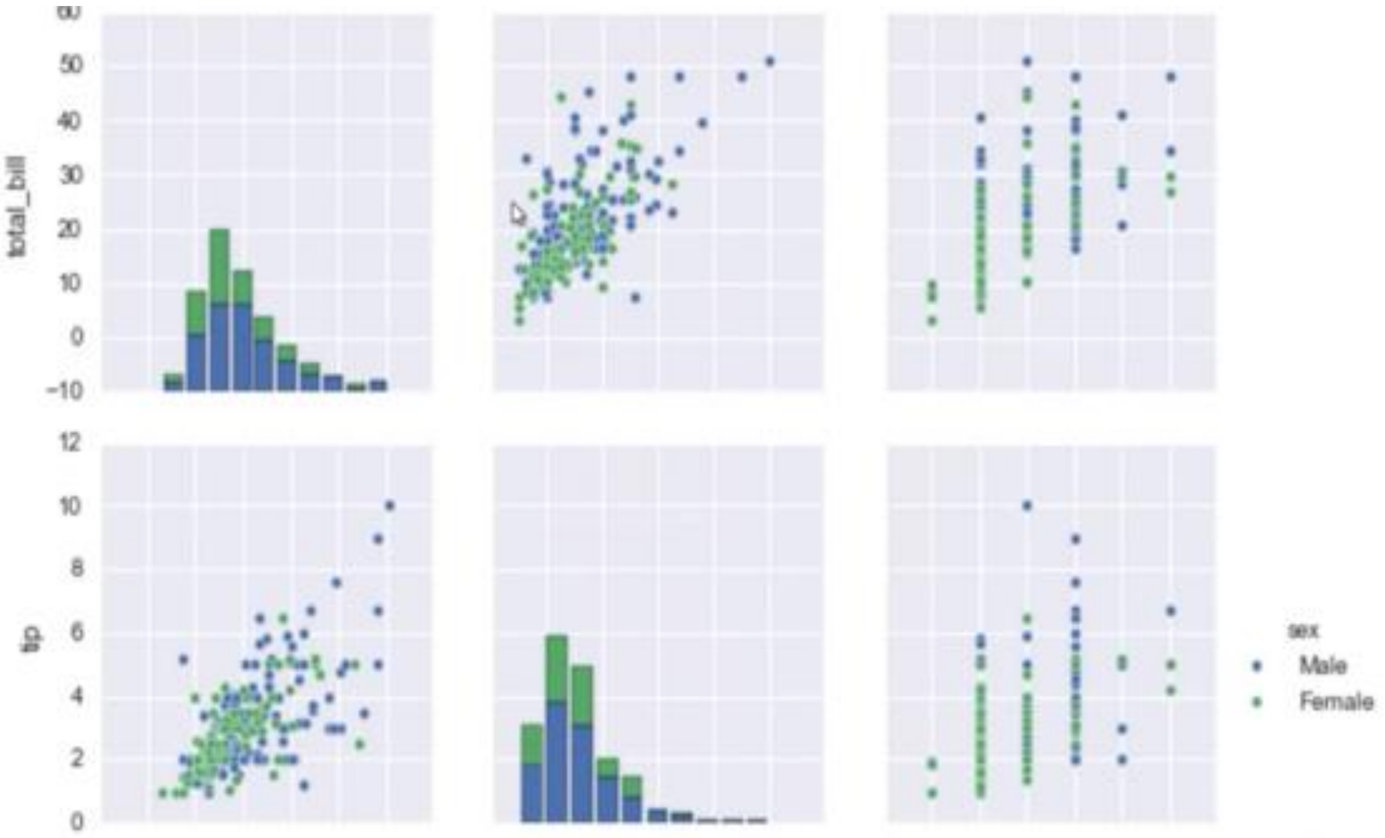
- Sns.pairplot(data=‘tips’ ,hue = ‘Sex’).

- Pair plot diagram is shown below.

- You can clearly observe that in below plot Size vs Total_bill, Tip Vs size.
- Hue = Sex will show the different color in plots.


total_bill
tip
size
60
50
40
30
20
10
0
−10
12
10
8
6
4
2
0
7
6
5
4
3
2
1
0
−10 0 10 20 30 40 50 60
0 2 4 6 8 10 12
0 1 2 3 4 5 6 7
total_bill
tip
size

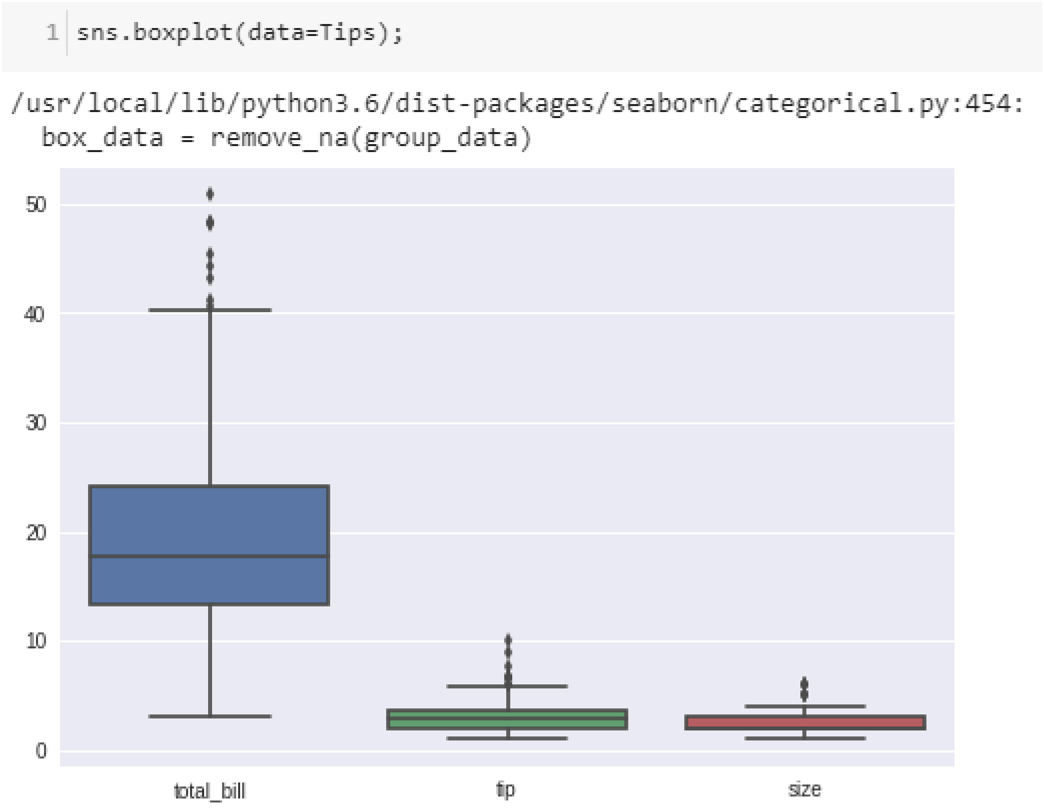
Pair plot with Hue = ‘sex’ option



Sex is one of the column in Tips table .so when you use Hue is equal to sex it will show separately

Box plots

- Box plots are essential visualization to understand how data is being distributed and find outliers.



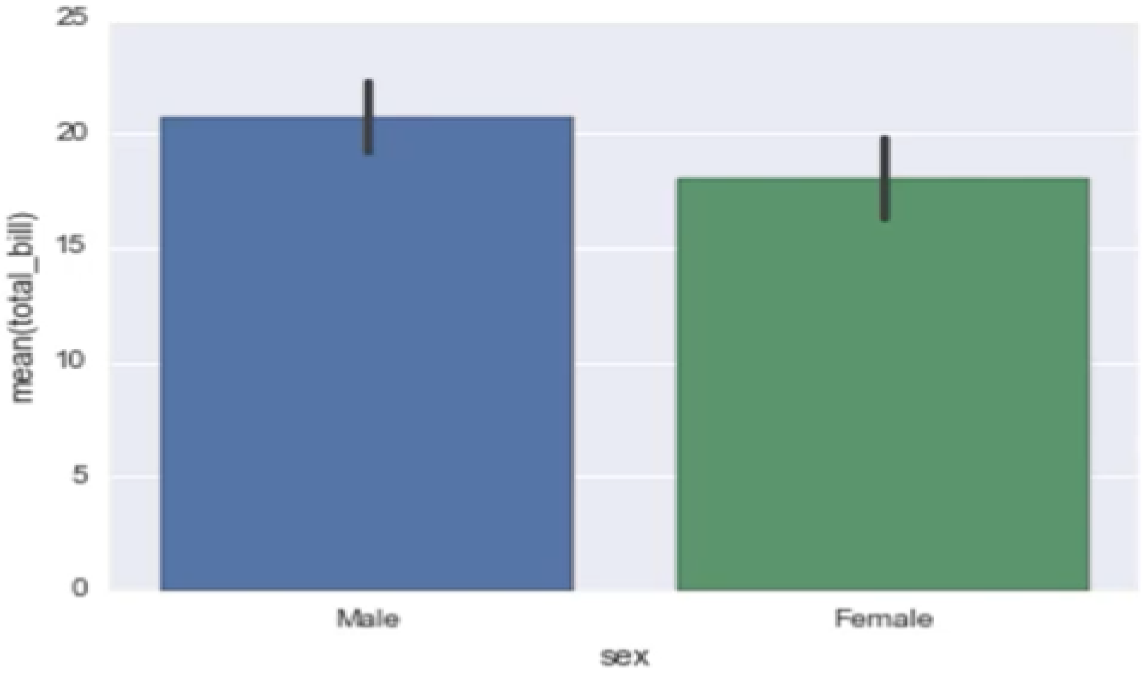
```
sns.boxplot(data=Tips);
```

```
/usr/local/lib/python3.6/dist-packages/seaborn/categorical.py:454:
  box_data = remove_na(group_data)
```

Categorical plots :

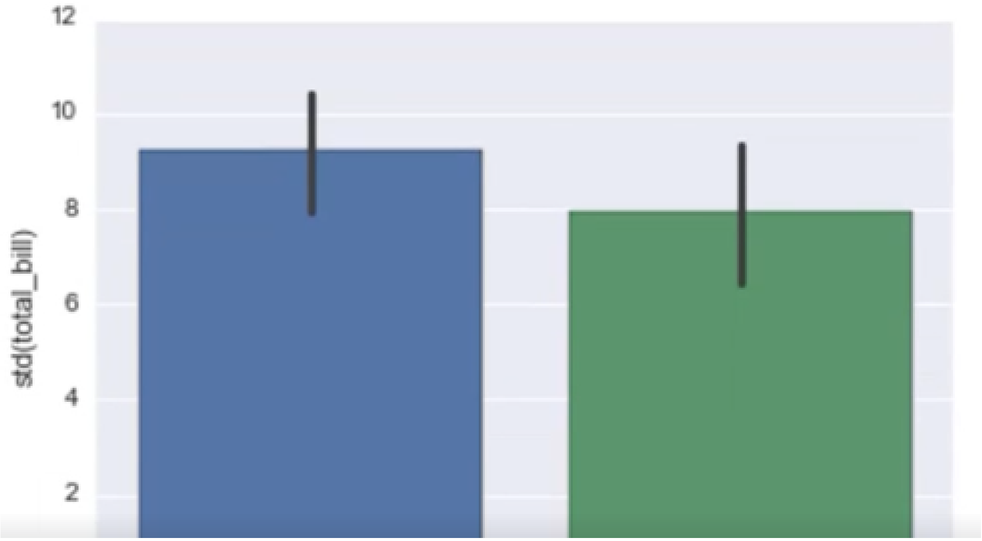
- Bar plot :

  sns.barplot(x= ‘sex’ ,y= ‘Total_bill’ , data= ‘tips’)
- This will give average bill for male and female.



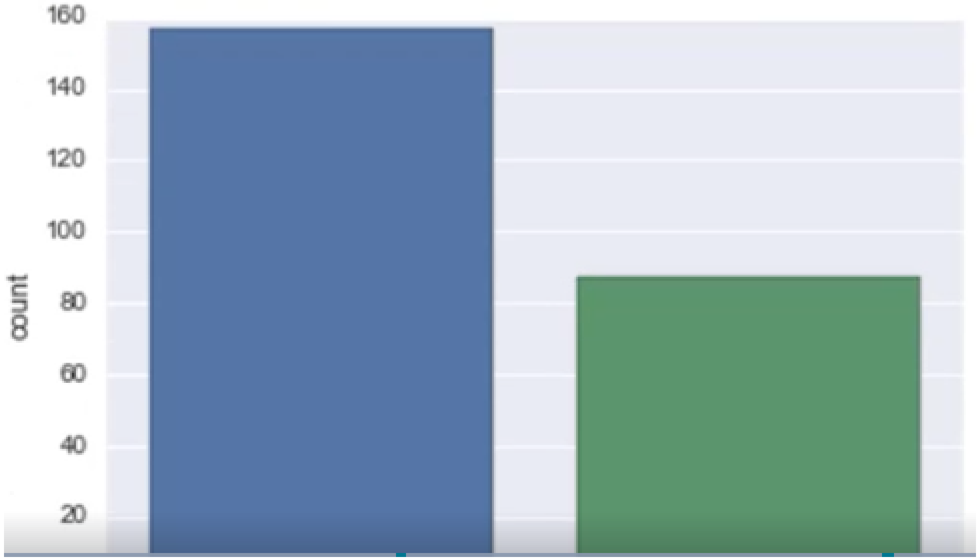
Bar plot with estimator

- In order to estimate the average , standrad devaition ,
- we need to import np library.
- Import numpy as np
- Sns.barplot(x=‘sex’,y=‘Total_bill’, data=‘tips’,estimator=np.std)



Count plot

- sns.countplot(x= ‘sex’ , data= ‘Tips’)
- Count plot will give us the total number in the data set.



- Thank You