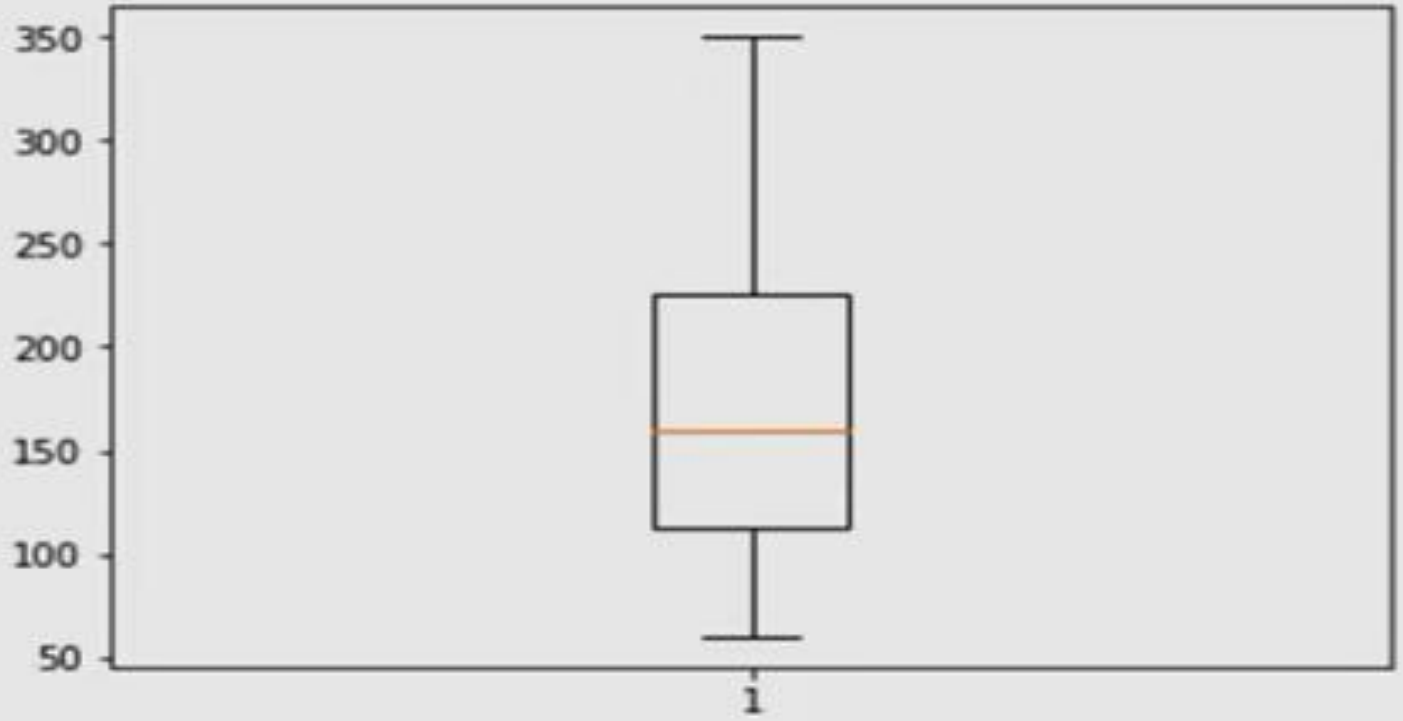
WEKA DATA MINING

Some statistics calculations, Creating box plot, scatter plot

BOX PLOT description

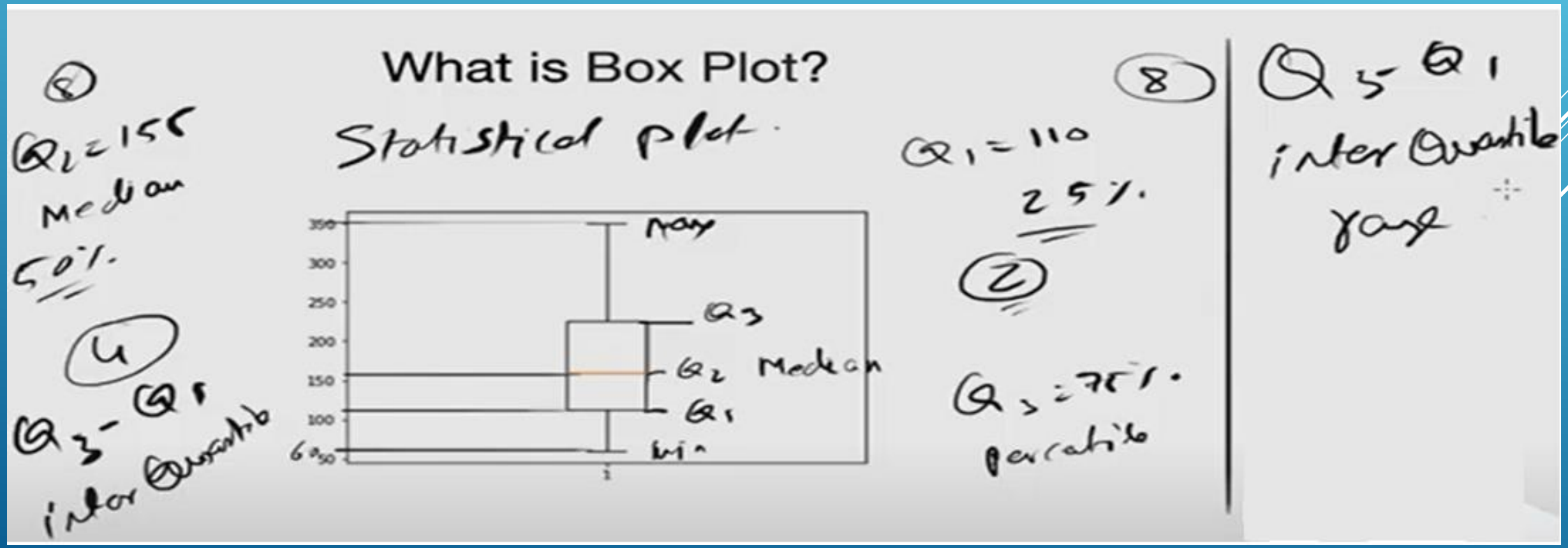
```
In [3]: sal_org_a = [200,300,150,90,120,350,60,170] # USD thousands per annum
```

```
In [2]: import matplotlib.pyplot as plt
        %matplotlib inline
        plt.boxplot(sal_org_a);
```



BOX PLOT description

Box plot description for salary in an organization *show in previous slide*.



SCATTER PLOT description

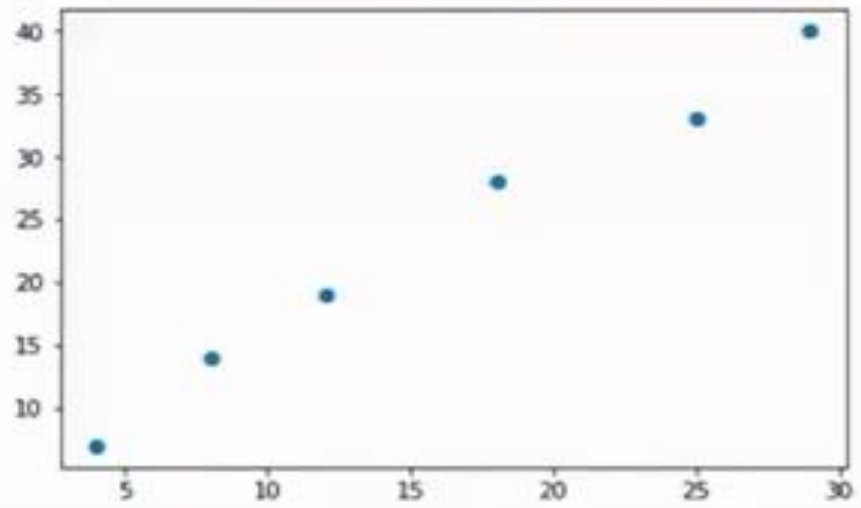
Scatter lots are used to understand relationship between two variables. Draw a *scatter plot* based on the two variables.

```
In [1]: import matplotlib.pyplot as plt
        %matplotlib inline

In [2]: x=[4,8,12,18,25,29]
        y=[7,14,19,28,33,40]

In [3]: plt.scatter(x,y)

Out[3]: <matplotlib.collections.PathCollection at 0x21e33c6b1d0>
```



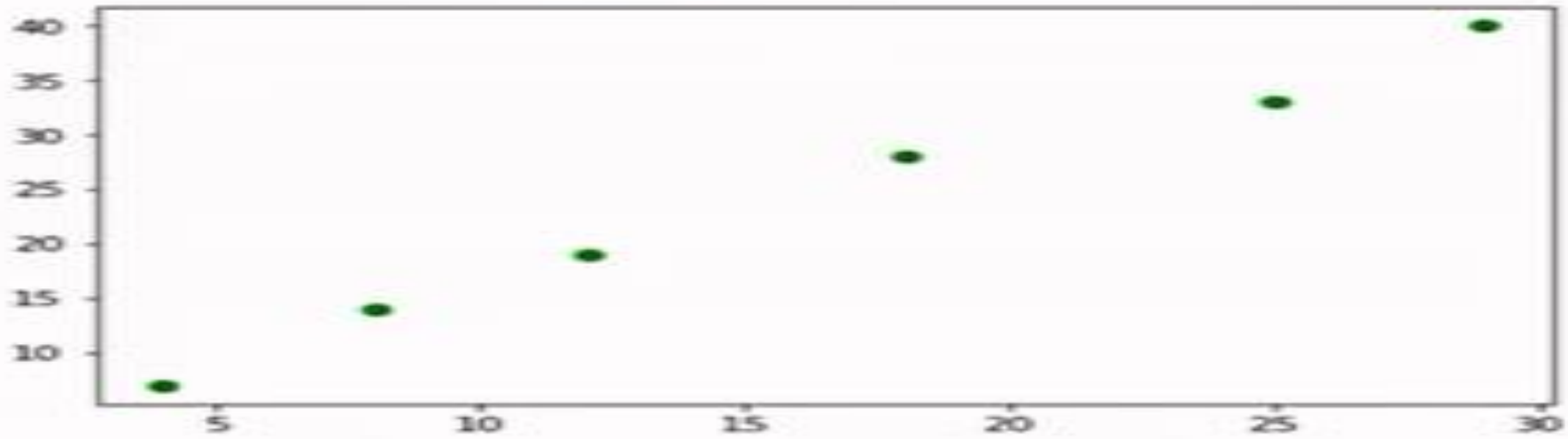
SCATTER Plot

Scatter lots are used to understand relationship between two variables. Draw a *scatter plot* based on the two variables.

```
In [1]: import matplotlib.pyplot as plt
        %matplotlib inline

In [2]: x=[4,8,12,18,25,29]
        y=[7,14,19,28,33,40]

In [4]: plt.scatter(x,y,color='
Out[4]: <matplotlib.collections.PathCollection at 0x21e3417a978>
```



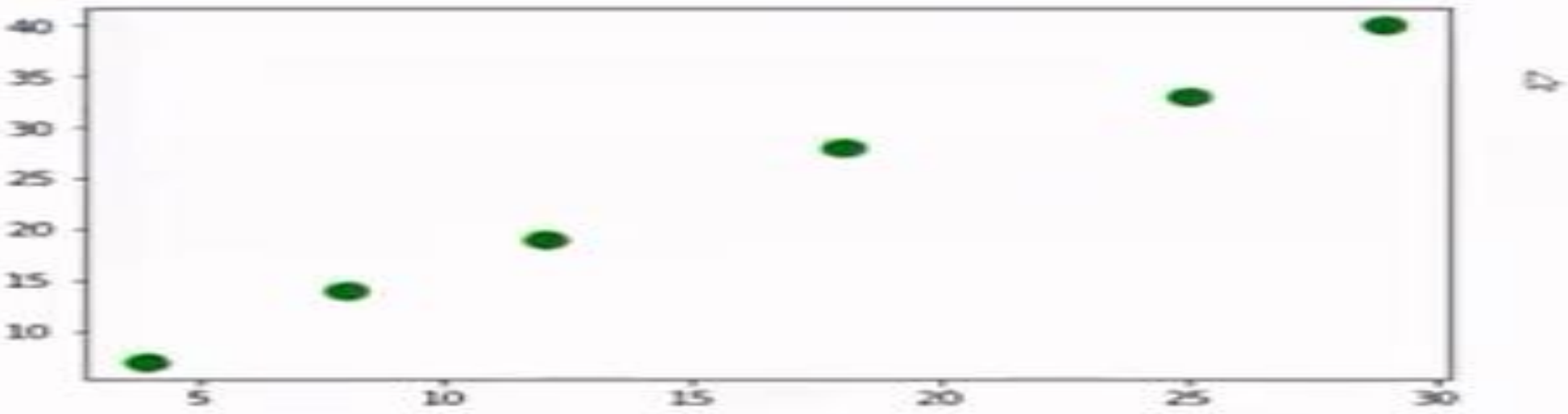
SCATTER Plot

Scatter lots are used to understand relationship between two variables. Draw a *scatter plot* based on the two variables.

```
In [1]: import matplotlib.pyplot as plt
        %matplotlib inline

In [2]: x=[4,8,12,18,25,29]
        y=[7,14,19,28,33,40]

In [5]: plt.scatter(x,y,color='g',s=100)
Out[5]: <matplotlib.collections.PathCollection at 0x21e33b2d390>
```



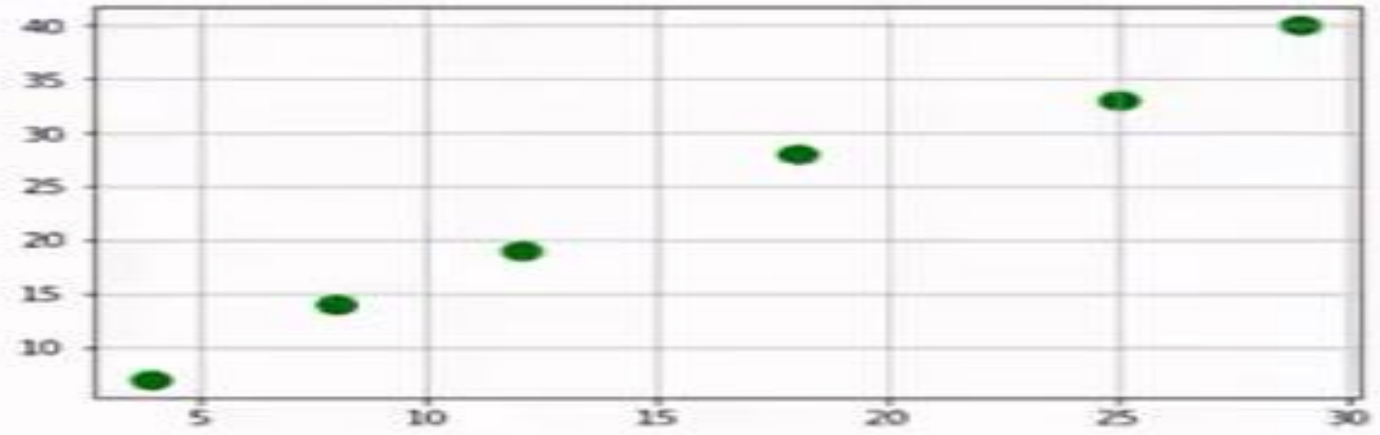
SCATTER Plot

Scatter lots are used to understand relationship between two variables. Draw a *scatter plot* based on the two variables.

```
In [1]: import matplotlib.pyplot as plt
        %matplotlib inline

In [2]: x=[4,8,12,18,25,29]
        y=[7,14,19,28,33,40]

In [6]: plt.scatter(x,y,color='g',s=100)
        plt.grid()
```



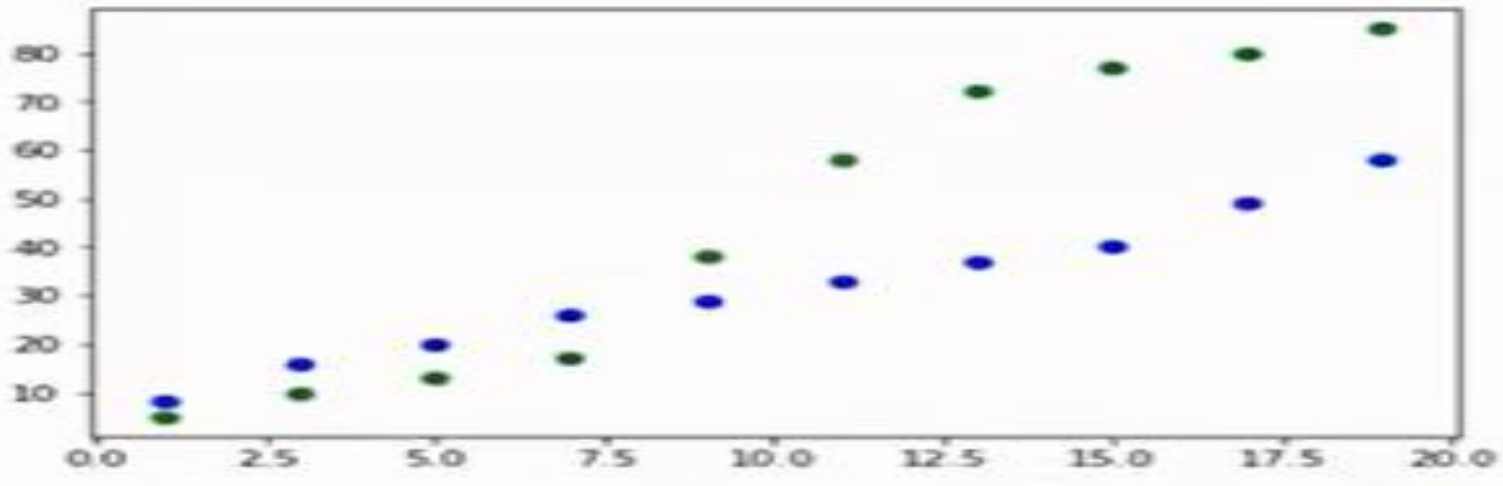
SCATTER Plot

Scatter lots are used to understand relationship between two variables. Draw a *scatter plot* based on the multiple variables.

```
In [7]: x1=[1,3,5,7,9,11,13,15,17,19]
        y1=[5,10,13,17,38,58,72,77,80,85]
        x2=[1,3,5,7,9,11,13,15,17,19]
        y2=[8,16,20,26,29,33,37,40,49,58]

In [8]: plt.scatter(x1,y1,color='g')
        plt.scatter(x2,y2,color='b')

Out[8]: <matplotlib.collections.PathCollection at 0x21e341ce390>
```

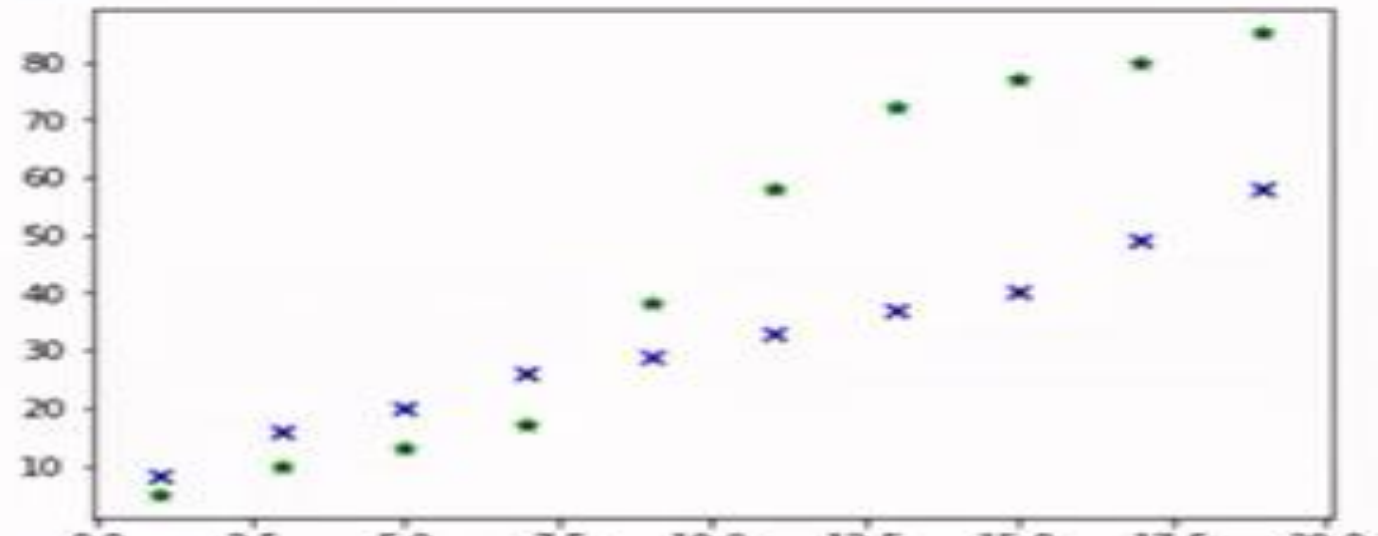


SCATTER Plot

Scatter lots are used to understand relationship between two variables. Draw a *scatter plot* based on the multiple variables.

```
In [9]: plt.scatter(x1,y1,color='g',marker='*')
        plt.scatter(x2,y2,color='b',marker='x')

Out[9]: <matplotlib.collections.PathCollection at 0x21e33da9470>
```

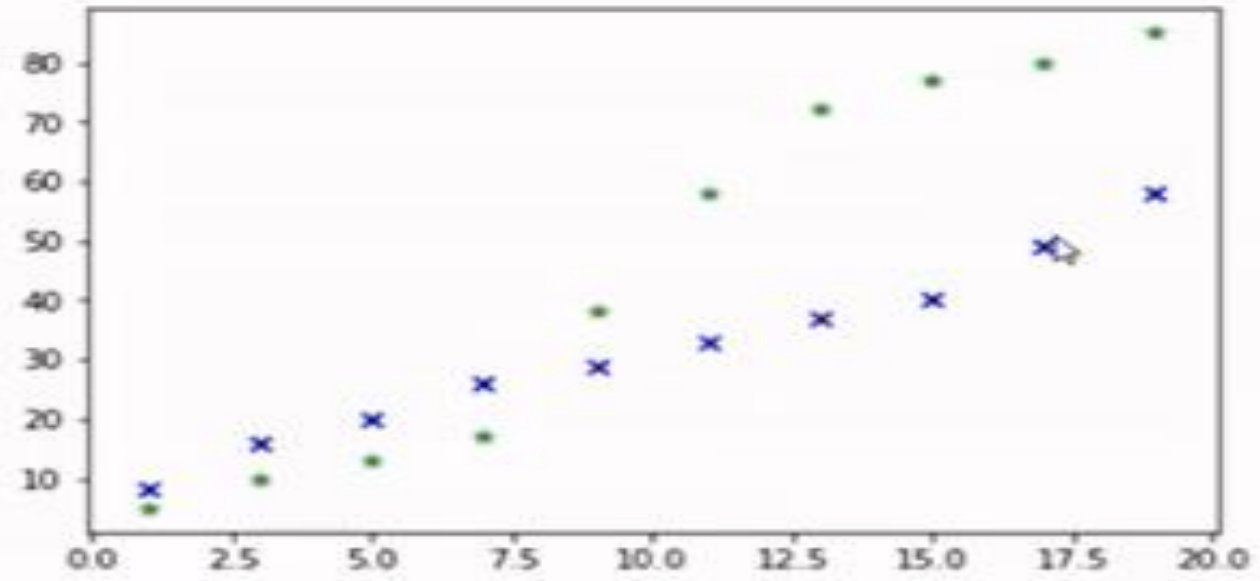


SCATTER Plot

Scatter lots are used to understand relationship between two variables. Draw a *scatter plot* based on the multiple variables.

```
In [11]: plt.scatter(x1,y1,color='g',marker='*',alpha=0.6)
         plt.scatter(x2,y2,color='b',marker='x',alpha=1)

Out[11]: <matplotlib.collections.PathCollection at 0x21e340e3550>
```

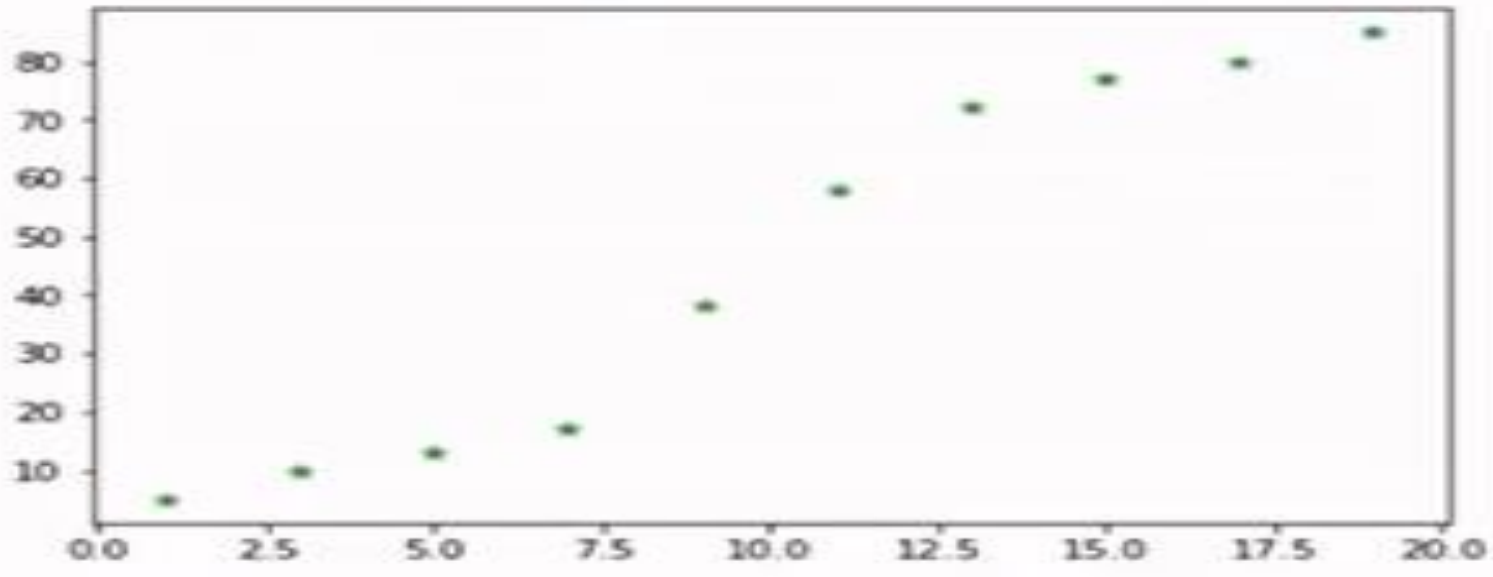


SCATTER Plot

Scatter lots are used to understand relationship between two variables. Draw a *scatter plot* based on the multiple variables.

```
In [12]: plt.scatter(x1,y1,color='g',marker='*',alpha=0.6)
         plt.scatter(x2,y2,color='b',marker='x',alpha=0.
```

```
Out[12]: <matplotlib.collections.PathCollection at 0x21e33cf2d68>
```

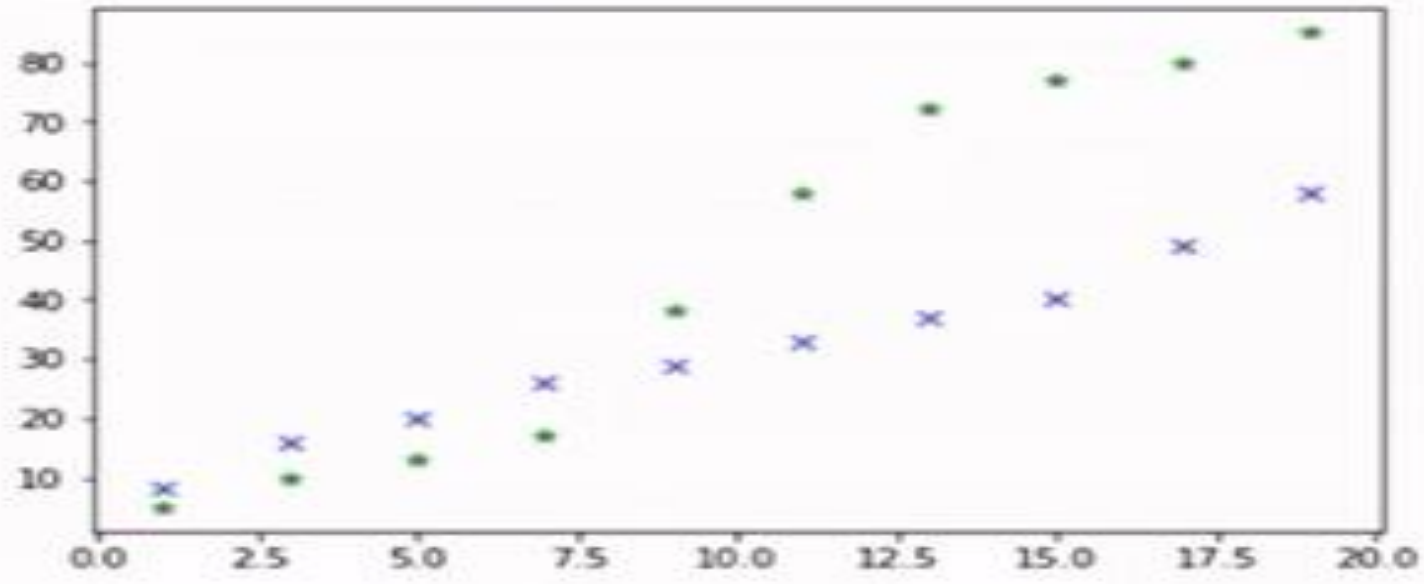


SCATTER Plot

Scatter lots are used to understand relationship between two variables. Draw a *scatter plot* based on the multiple variables.

```
In [13]: plt.scatter(x1,y1,color='g',marker='*',alpha=0.6)
         plt.scatter(x2,y2,color='b',marker='x',alpha=0.6)

Out[13]: <matplotlib.collections.PathCollection at 0x21e33d15e10>
```

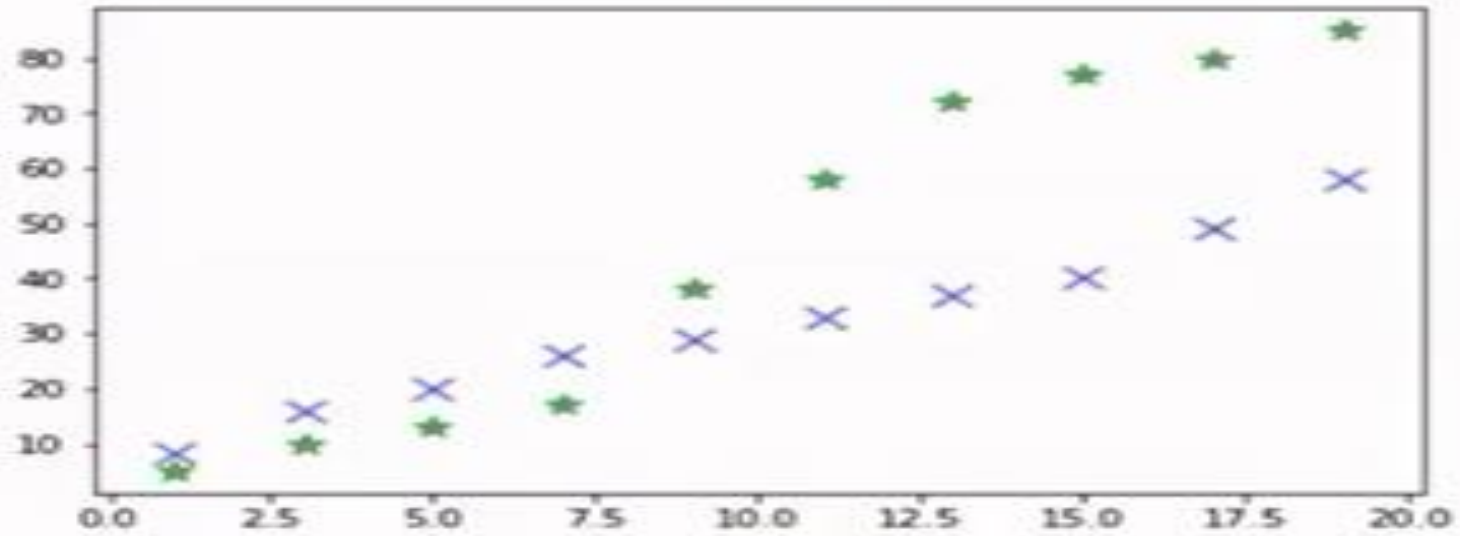


SCATTER Plot

Scatter lots are used to understand relationship between two variables. Draw a *scatter plot* based on the multiple variables.

```
In [14]: plt.scatter(x1,y1,color='g',marker='*',alpha=0.6,s=100)
         plt.scatter(x2,y2,color='b',marker='x',alpha=0.6,s=100)

Out[14]: <matplotlib.collections.PathCollection at 0x21e34518c18>
```

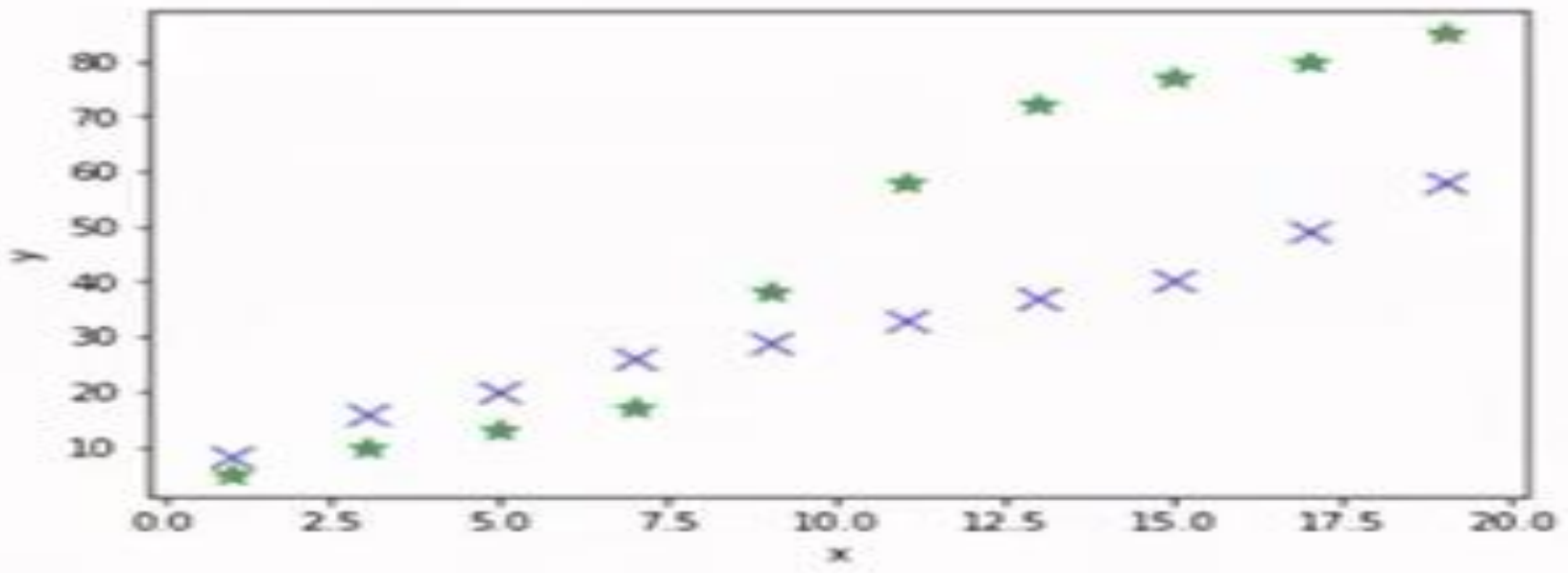


SCATTER Plot

Scatter lots are used to understand relationship between two variables. Draw a *scatter plot* based on the multiple variables.

```
In [15]: plt.scatter(x1,y1,color='g',marker='*',alpha=0.6,s=100)
         plt.scatter(x2,y2,color='b',marker='x',alpha=0.6,s=100)
         plt.xlabel('x')
         plt.ylabel('y')

Out[15]: Text(0, 0.5, 'y')
```

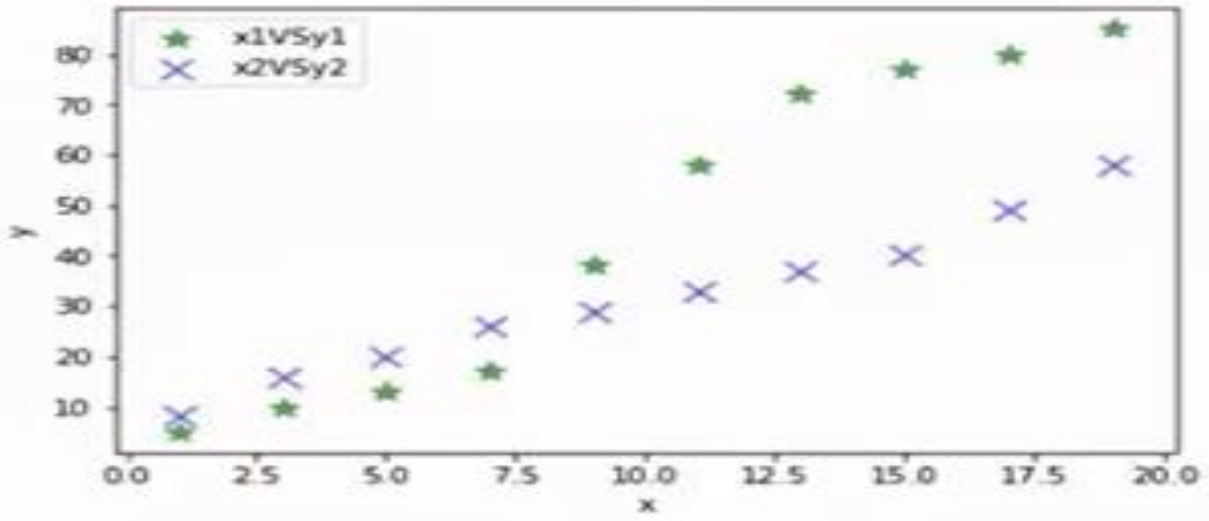


SCATTER Plot

Scatter lots are used to understand relationship between two variables. Draw a *scatter plot* based on the multiple variables.

```
In [16]: plt.scatter(x1,y1,color='g',marker='*',alpha=0.6,s=100,label='x1VSy1')
         plt.scatter(x2,y2,color='b',marker='x',alpha=0.6,s=100,label='x2VSy2')
         plt.xlabel('x')
         plt.ylabel('y')
         plt.legend()

Out[16]: <matplotlib.legend.Legend at 0x21e345c12e8>
```



SCATTER Plot

Scatter lots are used to understand relationship between two variables. Draw a *scatter plot* based on the multiple variables.

```
In [17]: plt.scatter(x1,y1,color='g',marker='*',alpha=0.6,s=100,label='x1VSy1')
         plt.scatter(x2,y2,color='b',marker='x',alpha=0.6,s=100,label='x2VSy2')
         plt.xlabel('x')
         plt.ylabel('y')
         plt.legend()
         plt.grid()
```