Draw histogram in python using matplotlib

1. What is histogram:-
   * was first introduced by Karl Pearson.
   * It is graphical representation which organizes a group of data points into user-specific ranges.
   * A histogram shows the frequency on the vertical axis and bins in the horizontal axis.
   * The "[bin](https://en.wikipedia.org/wiki/Data_binning)" (or "[bucket](https://en.wikipedia.org/wiki/Data_binning)") is the range of values—that is, divide the entire range of values into a series of intervals—and then count how many values fall into each interval.

**Example to understand histogram:-**

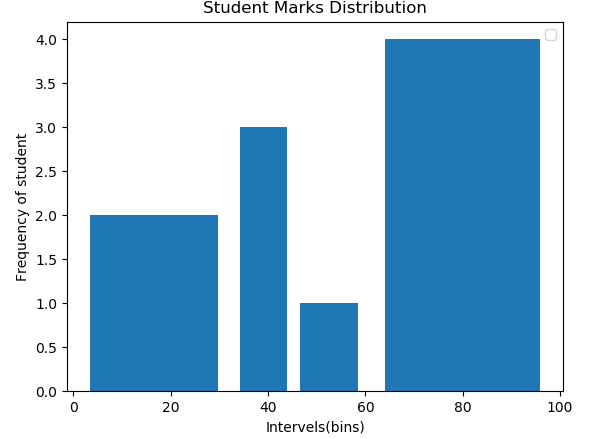
**Let students marks are given:-**

|  |  |  |
| --- | --- | --- |
| s.no | student name | marks |
| 1 | amit | 40 |
| 2 | dinesh | 60 |
| 3 | pragya | 55 |
| 4 | abhishek | 20 |
| 5 | rita | 35 |
| 6 | shyam | 70 |
| 7 | noman | 60 |
| 8 | piyush | 89 |
| 9 | samrth | 20 |
| 10 | tanmay | 33 |

**Find the Number of student in following Range:-**

|  |  |  |
| --- | --- | --- |
| 0-33 | >0 and <33 | 2 |
| 33-45 | >=33 and <45 | 3 |
| 45-60 | >=45 and <60 | 1 |
| 60-100 | >=60 | 4 |

**How histrogram help to visualize the result from above observation:-**

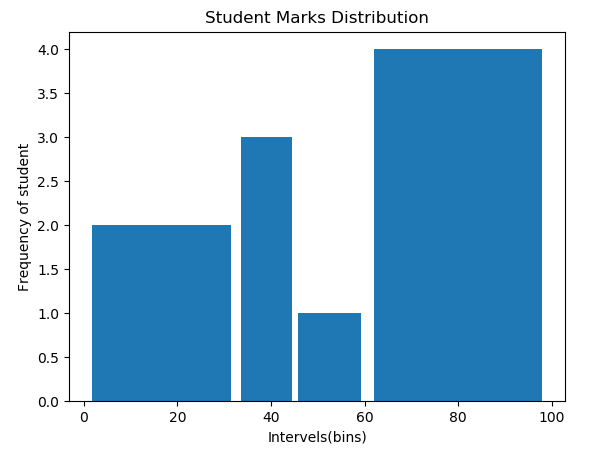


**How to draw in python:-**

Function:- matplotlib.pyplot.hist()

|  |  |
| --- | --- |
| **Type of Plot** | **Function to be used** |
| histogram | **plt.hist()** |

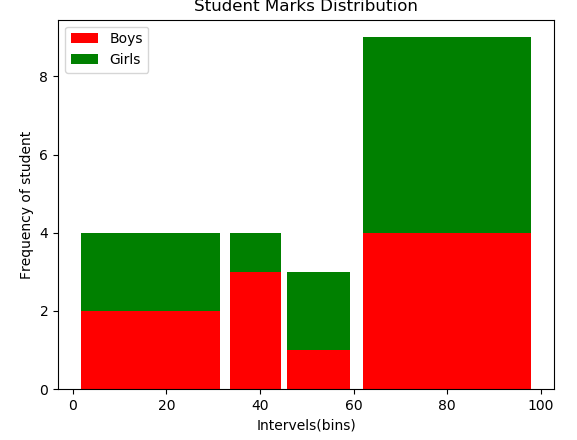
|  |  |  |
| --- | --- | --- |
| **Important parameter of hist() function:-**  matplotlib.pyplot.hist(x, bins=None, weights=None,histtype='bar',orientation='vert ical', rwidth=None, color=None, label=None, stacked=False) | | |
| **x** | array or sequence of (n,) arrays | |
| **bins** | int or sequence or str, optional (default: 10) if *bins* is:  [1,2,3,4]  then it have 3 bins:-  first bin is [1, 2) (including 1, but excluding 2)  the second [2, 3).  And last bin, however, is [3, 4], which *includes* 4. | |
| **Drawing simple histogram (Taking default bins=10)** | import matplotlib.pyplot as plt  **marks=[40,60,55,20,35,70,60,89,20,33]**  **plt.hist(marks)** plt.xlabel("Intervels(bins)") plt.ylabel("Frequency of student") plt.title("Student Marks Distribution") plt.show() |  |
| **Drawing simple histogram (Taking bins=5)** | import matplotlib.pyplot as plt marks=[40,60,55,20,35,70,60,89,20,33]  **plt.hist(marks,bins=5)** plt.xlabel("Intervels(bins)") plt.ylabel("Frequency of student") plt.title("Student Marks Distribution") plt.show() |  |
| **Finding values of bins created by function** | import matplotlib.pyplot as plt marks=[40,60,55,20,35,70,60,89,20,33]  **a=plt.hist(marks) print(a)** | |



|  |  |  |
| --- | --- | --- |
|  | **a,b,c = plt.hist(marks) print(a)**  **print(b) print(c)**     * hist() function **Returns: n**array or list of arrays, **bins** array, **patches** list or list of lists | |
| **Drawing** | import matplotlib.pyplot as plt |  |
| **simple histogram (Taking user defined 3 bins)** | marks=[40,60,55,20,35,70,60,89,20,33]  **plt.hist(boy,bins=[0,33,45,60,100])** plt.xlabel("Intervels(bins)") plt.ylabel("Frequency of student") plt.title("Student Marks Distribution")  plt.legend() |
|  | plt.show() |
|  | **Creates 4 bins:-** |
|  | **1. >0 to <33** |
|  | **2. >=33 to <45** |
|  | **3. >=45 to <60** |
|  | **4. >=60 to <=100** |
| **Using rwidth parameter (**The relative width of the bars as a fraction of the bin width) | import matplotlib.pyplot as plt marks=[40,60,55,20,35,70,60,89,20,33]  plt.hist(marks,bins=[0,33,45,60,100],**rwidt h=0.9)**  plt.xlabel("Intervels(bins)") plt.ylabel("Frequency of student") plt.title("Student Marks Distribution") plt.show() |  |

|  |  |  |
| --- | --- | --- |
| **Using color** | import matplotlib.pyplot as plt marks=[40,60,55,20,35,70,60,89,20,33]  plt.hist(marks,bins=[0,33,45,60,100],rwidt h=0.9,**color=”green”)** plt.xlabel("Intervels(bins)") plt.ylabel("Frequency of student") plt.title("Student Marks Distribution") plt.show() |  |
| **Using weights** | import matplotlib.pyplot as plt marks=[40,60,55,20,35,70,60,89,20,33]  plt.hist(marks,bins=[0,33,45,60,100],**weigh ts=[40,60,55,20,35,70,60,89,20,33]**)  plt.xlabel("Intervels(bins)") plt.ylabel("Frequency of student") plt.title("Student Marks Distribution") plt.show()  **Explanation:-**  Each value in *x* only contributes its associated weight towards the bin count |  |
|  | plt.hist(marks,bins=[0,33,45,60,100],**weigh ts=[10,10,10,10,10,10,10,10,10,10]**) |  |

|  |  |  |
| --- | --- | --- |
| **Orientation**  {'horizontal', 'vertical'}, optional | import matplotlib.pyplot as plt marks=[40,60,55,20,35,70,60,89,20,33]  plt.hist(marks,bins=[0,33,45,60,100],rwidt h=0.9,**orientation="horizontal"**) plt.xlabel("Intervels(bins)") plt.ylabel("Frequency of student") plt.title("Student Marks Distribution") plt.show() |  |
| **Using histtype** 'bar'(Default)  ,  'barstacked', 'step', 'stepfilled', optional | import matplotlib.pyplot as plt marks=[40,60,55,20,35,70,60,89,20,33]  plt.hist(marks,bins=[0,33,45,60,100],**histt ype="step"**,rwidth=0.9) plt.xlabel("Intervels(bins)") plt.ylabel("Frequency of student") plt.title("Student Marks Distribution") plt.show()   * *Ignoring rwidth parameter* |  |
|  | import matplotlib.pyplot as plt marks=[40,60,55,20,35,70,60,89,20,33]  plt.hist(marks,bins=[0,33,45,60,100],**histt ype="stepfilled"**) plt.xlabel("Intervels(bins)") plt.ylabel("Frequency of student") plt.title("Student Marks Distribution") plt.show() |  |
| **Display multiple distribution frequency together** | import matplotlib.pyplot as plt **boys=[40,60,55,20,35,70,60,89,20,33] girls=[50,68,44,90,32,15,86,74,54,68]**  **plt.hist([boys,girls],bins=[0,33,45,60,100**  **],color=["red","green"],rwidth=0.9)** plt.xlabel("Intervels(bins)") plt.ylabel("Frequency of student") plt.title("Student Marks Distribution") plt.show() |  |



|  |  |  |
| --- | --- | --- |
| **Using Label** | import matplotlib.pyplot as plt boys=[40,60,55,20,35,70,60,89,20,33] girls=[50,68,44,90,32,15,86,74,54,68]  a=plt.hist([boys,girls],bins=[0,33,45,60,1 00**],label=["Boys","Girls"],**color=["red","g reen"],rwidth=0.9) plt.xlabel("Intervels(bins)") plt.ylabel("Frequency of student") plt.title("Student Marks Distribution") **plt.legend()**  plt.show() |  |
| **Using stacked** | If True, multiple data are stacked on top of each other  If False multiple data are arranged side by side if histtype is 'bar' or on top of each other if histtype is 'step'  import matplotlib.pyplot as plt boys=[40,60,55,20,35,70,60,89,20,33] girls=[50,68,44,90,32,15,86,74,54,68]  plt.hist([boys,girls],bins=[0,33,45,60,100  ],**stacked=True**,label=["Boys","Girls"],colo r=["red","green"],rwidth=0.9) plt.xlabel("Intervels(bins)") plt.ylabel("Frequency of student") plt.title("Student Marks Distribution") plt.legend()  plt.show() |  |