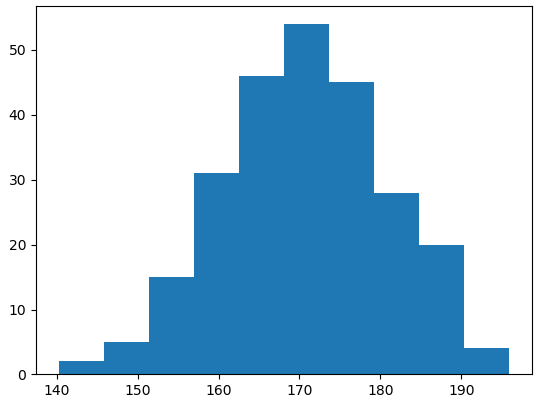
## Histogram

A histogram is a graph showing frequency distributions.

It is a graph showing the number of observations within each given interval.

Example: Say you ask for the height of 250 people, you might end up with a histogram like this:



You can read from the histogram that there are approximately:

2 people from 140 to 145cm  
5 people from 145 to 150cm  
15 people from 151 to 156cm  
31 people from 157 to 162cm  
46 people from 163 to 168cm  
53 people from 168 to 173cm  
45 people from 173 to 178cm  
28 people from 179 to 184cm  
21 people from 185 to 190cm  
4 people from 190 to 195cm

## Create Histogram

In Matplotlib, we use the hist() function to create histograms.

The hist() function will use an array of numbers to create a histogram, the array is sent into the function as an argument.

For simplicity we use NumPy to randomly generate an array with 250 values, where the values will concentrate around 170, and the standard deviation is 10.

import numpy as np  
  
x = np.random.normal(170, 10, 250)  
  
print(x)

This will generate a *random* of 250 values as result, and could look like this( your reference only not for exam)

[167.62255766 175.32495609 152.84661337 165.50264047 163.17457988

162.29867872 172.83638413 168.67303667 164.57361342 180.81120541

170.57782187 167.53075749 176.15356275 176.95378312 158.4125473

187.8842668 159.03730075 166.69284332, etc.,

Example

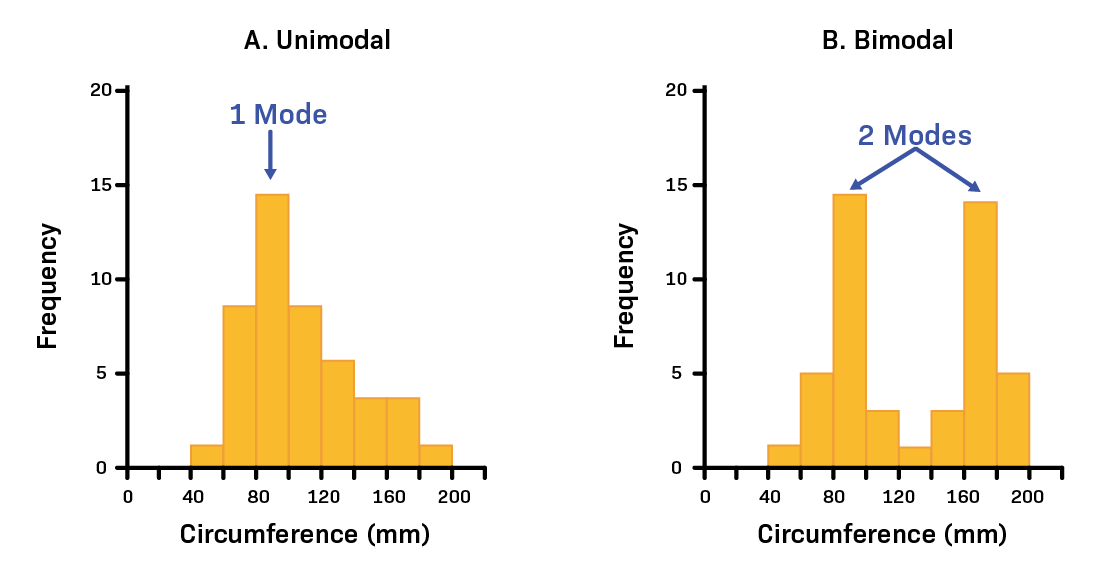
A simple histogram: the plt.hist() takes three arguments

first data: mean

second data: standard deviation and third data: no.of data

|  |  |
| --- | --- |
| import matplotlib.pyplot as plt  import numpy as np  x = np.random.normal(170, 10, 250)  plt.hist(170,10,250)  plt.show() |  |

* Histograms are one of the most intuitive ways of representing the shape of a data set's distribution along a single [numeric variable](https://www.labxchange.org/library/pathway/lx-pathway:3eec864d-746b-41dc-b3e4-8bb6a03949fe/items/lx-pb:3eec864d-746b-41dc-b3e4-8bb6a03949fe:html:b125ff34).
* Histograms allow us to quickly approximate center, spread, and shape. They also allow us to identify whether there are outliers.
* In particular, the shape of a histogram reflects whether the data is symmetric or skewed and whether there are one or multiple modes.
* When comparing 2 histograms, check that they are on the same scale and have the same bins.





**Measures of Center**

* For interval or ratio level data, one measure of center is the **mean**. The **population mean** is denoted by μ�, while the **sample mean** intended to estimate it is denoted by x¯¯¯�¯. Both values are calculated in a very similar way. Assuming the population has size N�, a sample has size n�, and x� spans across all available data values in the population or sample, as appropriate, we find these means by calculating



* The **median**, denoted by Q2 (or med) is the middle value of a data set when it is written in order. In the case of an even number of data values (and thus no exact middle), it is the average of the middle two data values. It is not affected by the presence of extreme values in the data set. Unlike the mean, it can sometimes† even suggest a central value for ordinal data.