# Measures of central tendency

Traditional statistics has 3 important parameters called:

* Mean
* Median
* Mode

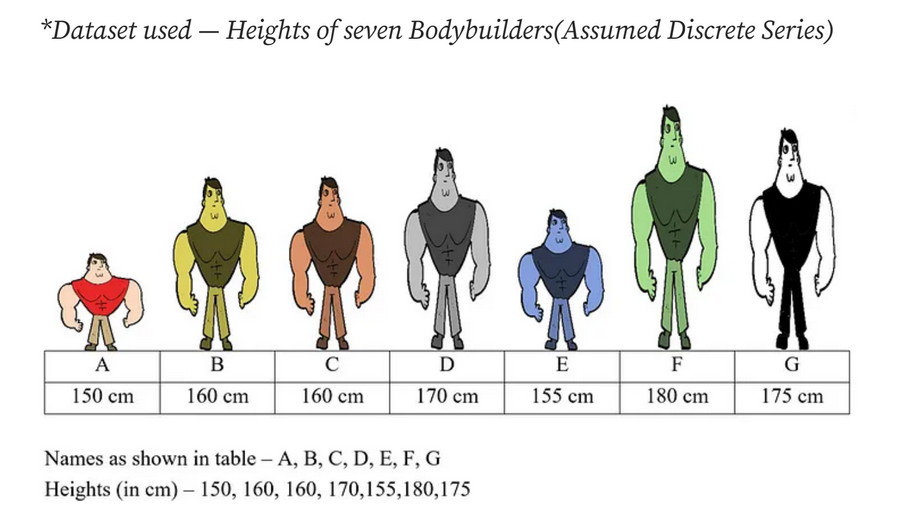
All these 3 refer to one single aspect called the central tendency. The idea of central tendency is that there may be one single value that can possibly describe the data to the best extent.

<https://drive.google.com/file/d/1Fe1BwmXLMSvq2czjZyUdCL0z_2N_FYUb/view?usp=drive_link>

# Mean, Median & Mode – Which central tendency measure to use & when?

To represent a dataset as a 1-number summary, we use central tendency measure.

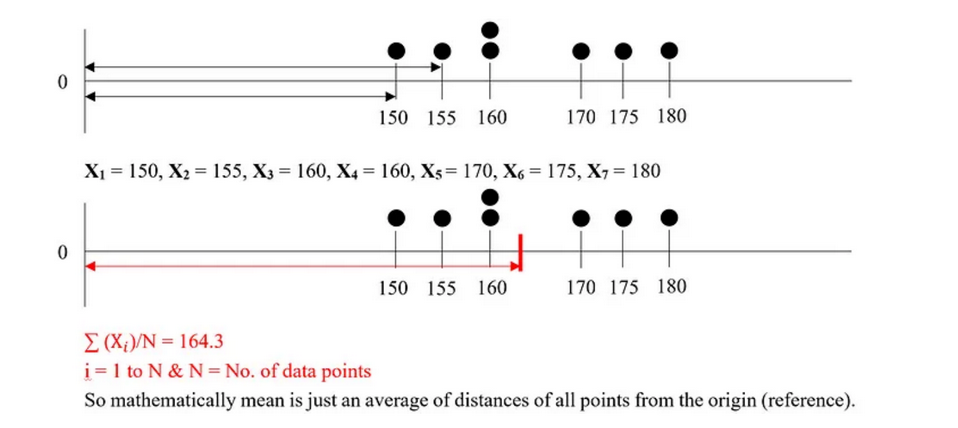
Why was there a need for 3 measures when only 1 (mean) could have done the job?

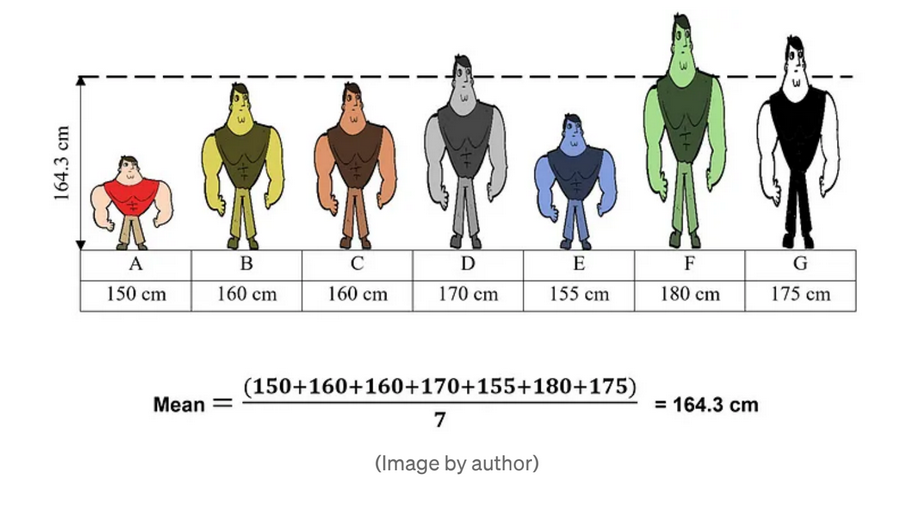


## Let’s start with the calculation of the Mean of this data:

150,160,160,170,155,180,175 – What do these numbers reflect?

If we put them on a number line, each point will be nothing but a distance from a reference point (in this case=0)

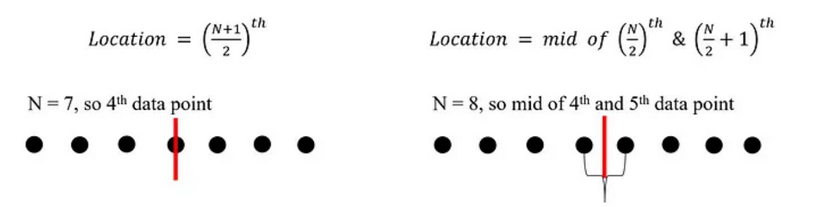


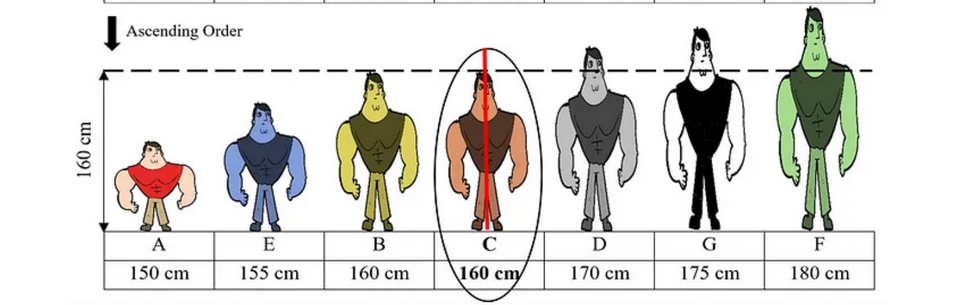


## Initiating calculation of the median of this data:

Steps:

1. Arrange the data points in an ascending order
2. Cross-section split, where the data points lie on the upper side and the remaining half on the lower side, is the median measure. Think of this as you are trying to partition the datapoints into 2-halves using a separator
3. If the data points count is odd, then there is one central value lying on the separator which is the median itself, otherwise it is the average of the 2 points lying on either side of the separator.





For the given data set of bodybuilders, N is odd and as is evident from the visual above, there are equal parts on either side of the median which is equal to 160 (it could also be B)

## Initiating calculation of the mode of this model:

This is the easiest one to calculate, just determine the frequency of the occurrence of each data point in the data, and the one with the highest frequency is the mode of the data. This measure can also be used when the data is non-numerical.

As there are two bodybuilders with height 160, this implies the mode of this dataset will be 160

# Now the most important part of this discussion

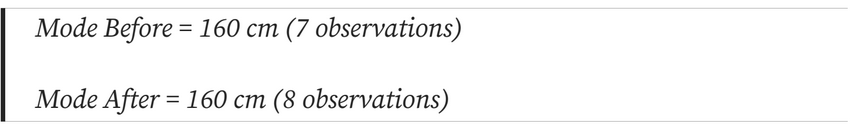
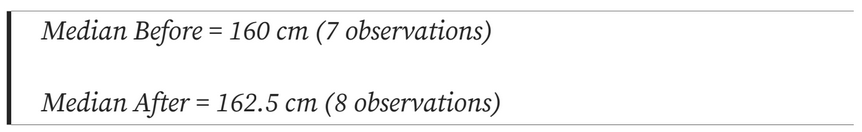
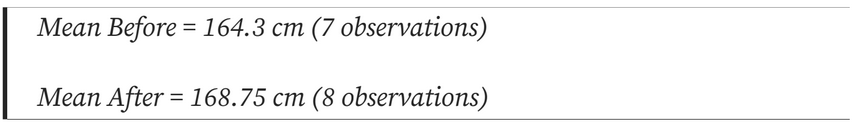
Why was there a need to have these 3 measures of central tendency instead of just one

To get a 1-number summary (central tendency), it is always intended to get an unbiased reflection of the whole data with that single measure.

However, we we’ll notice in the exercise below that sometimes Mean alone fails to stay unbiased and that causes our measure of central tendency to drift high or low

To illustrate this let’s add one more point of external data

1 more body builder (H) with a height=200cm



# Things to notice:

* Mean
  + Very sensitive to large additions to the dataset
* Median
  + Changes very little
* Mode
  + No change

These different measures are used to serve as the best feasible alternates for dealing with their innate biases (as shown above)