**Probability Distribution Functions:**

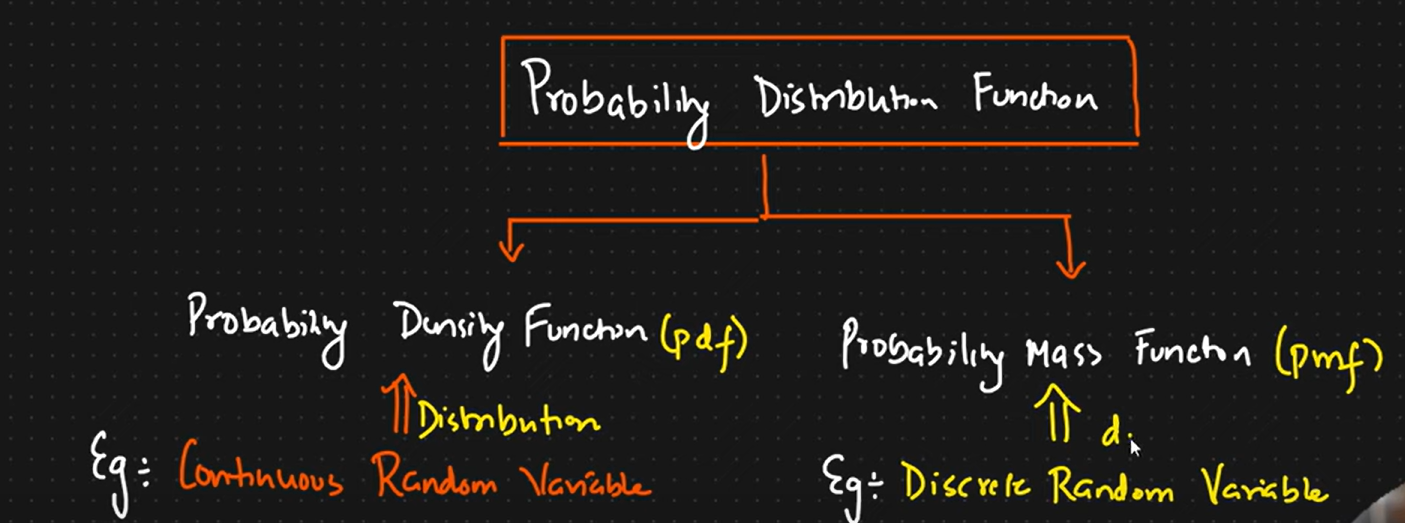
Aim: To understand different distributions of data

Definition: Probability distribution functions define how the probabilities are distributed over the values of a random variable.

Probability distribution functions are of two main types:

1. Probability Mass functions (PMF): It is used for discrete random variable.

2. Probability Density function (PDF): It is used for continuous random variable.

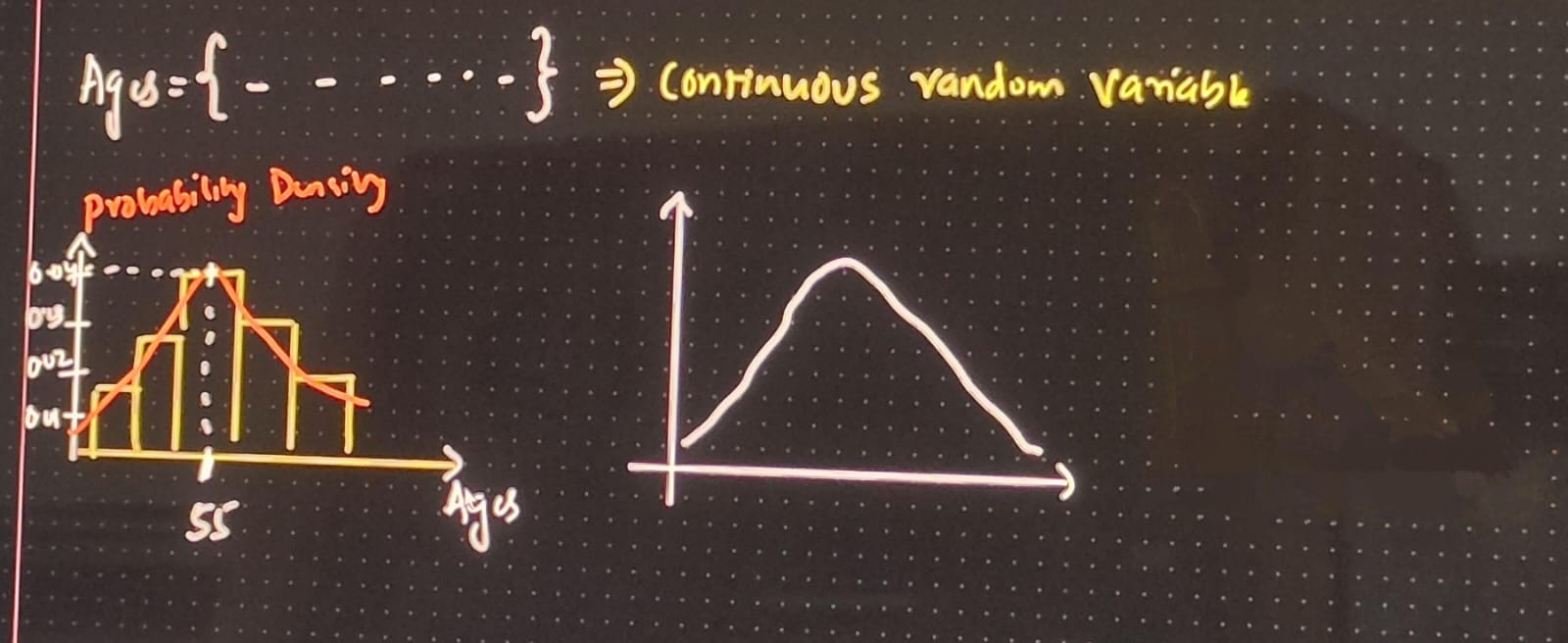


Example:  
Let’s say we have a random variable ages which is a set of continuous random variables.

If we construct a histogram with the y axis having frequencies and the x axis ages

If we smoothen the histogram, we get the distribution.

Once we smoothen, it, the y axis becomes the probability density.



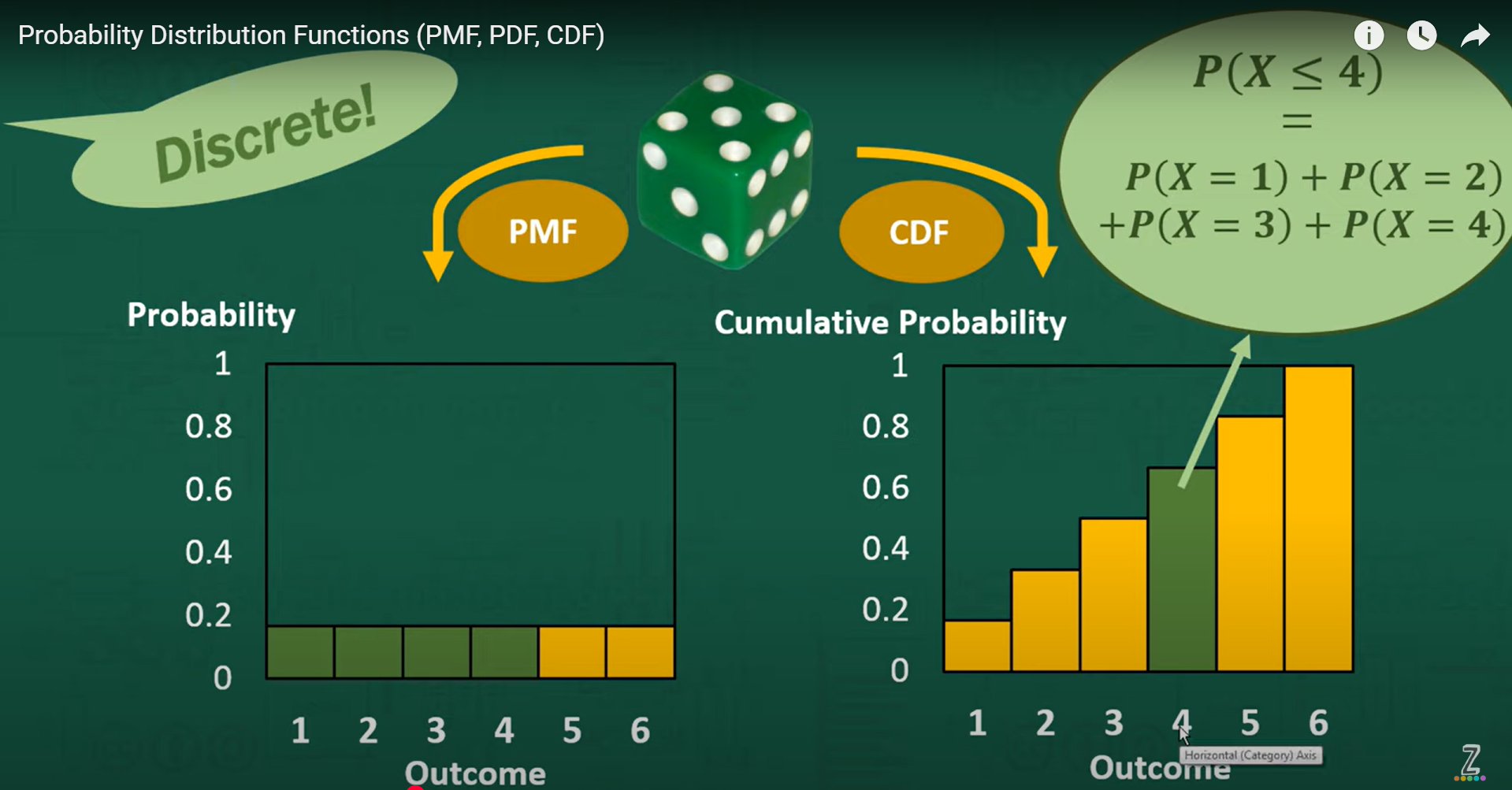
Probability mass function: (Discrete Random Variable)

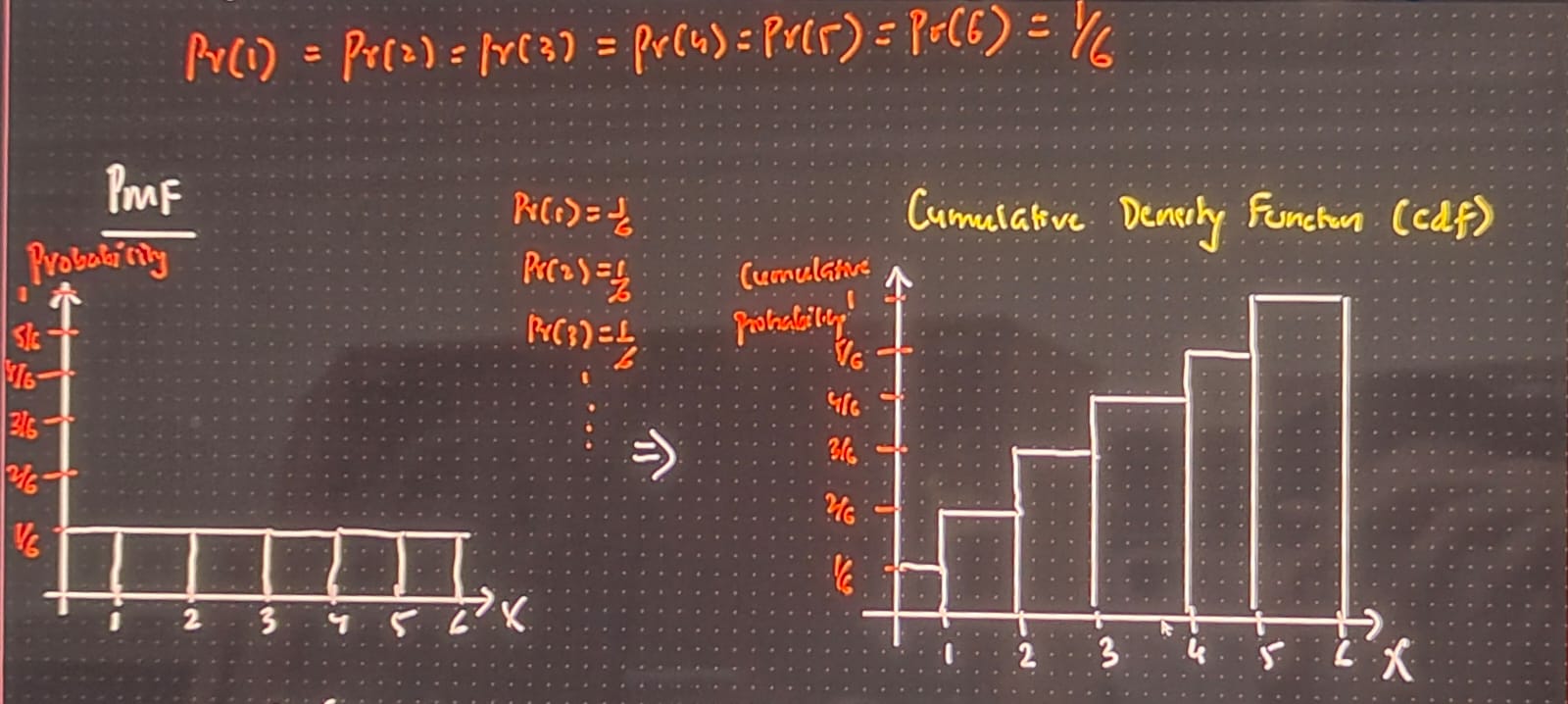
Eg: Rolling a fair dice [1,2,3,4,5,6]

P(1) = P(2) = P(3) = P(4) = P(5) = P(6) = 1/6

X axis will be the outcomes of rolling a dice.

Y axis will be probability values.





Cumulative density function (CDF):

X axis will be the outcomes of rolling a dice.

Y axis will be cumulative probability.

We keep on adding the probabilities as we go along.

We have to keep on combining the probabilities as we go from 1 to 6. (Start to end)

For example for 1 it would be 1/6.

Then for 2 it would be P(1&2) = 1/6 + 1/6 = 2/6

Then for 2 it would be P(1&2&3) = 1/6 + 1/6 + 1/6 = 3/6

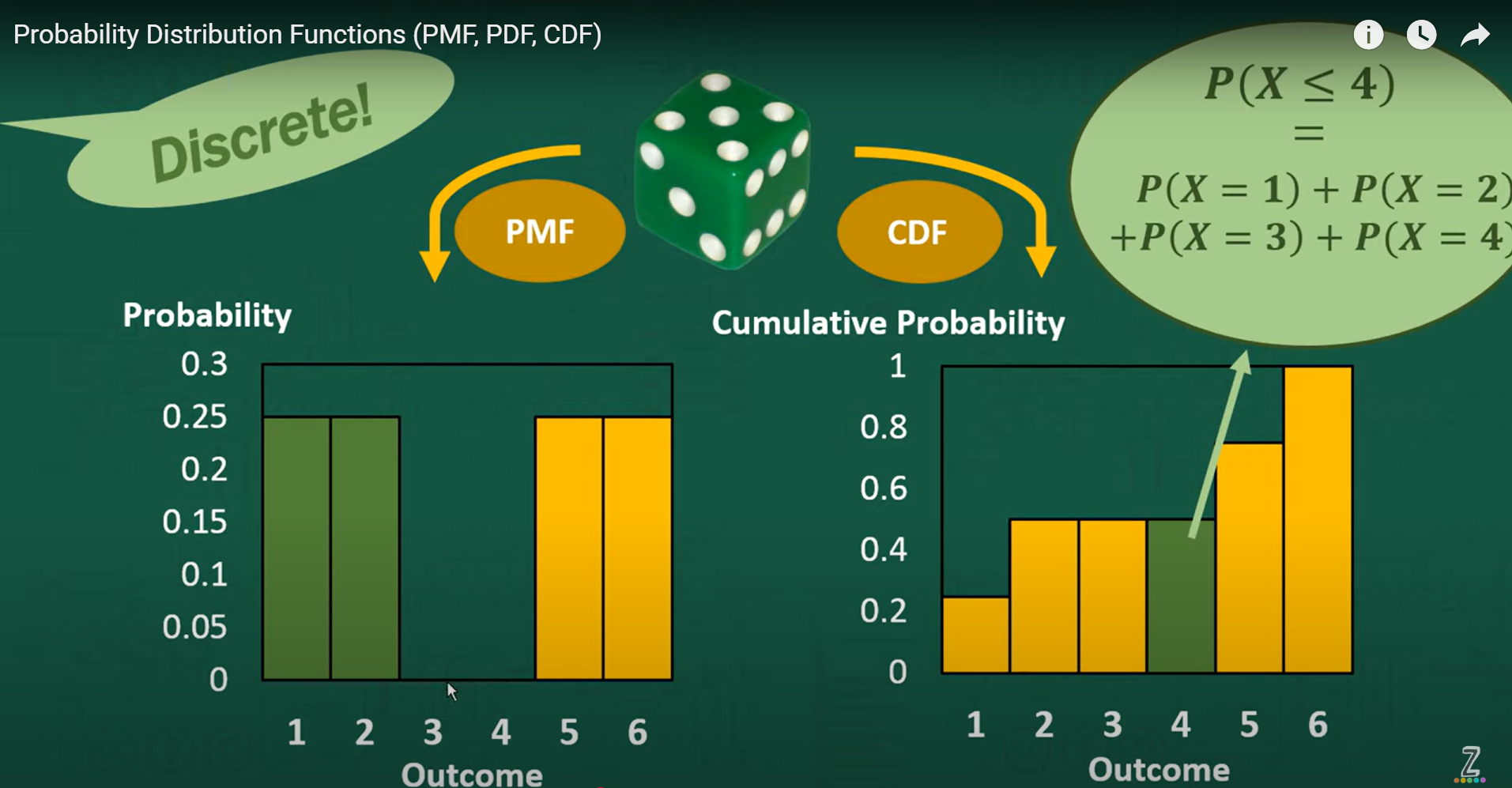
Keep adding till we get 1.

Example application:

We can find the P(X <= 2)

P(X<=2) = P(X=1) + P(X=2) = 2/6 = 1/3

If the dice could not roll 3 and 4.



Flatness in the CDF indicates, there is no mass for 3 and 4.

Probability Density Function (PDF):

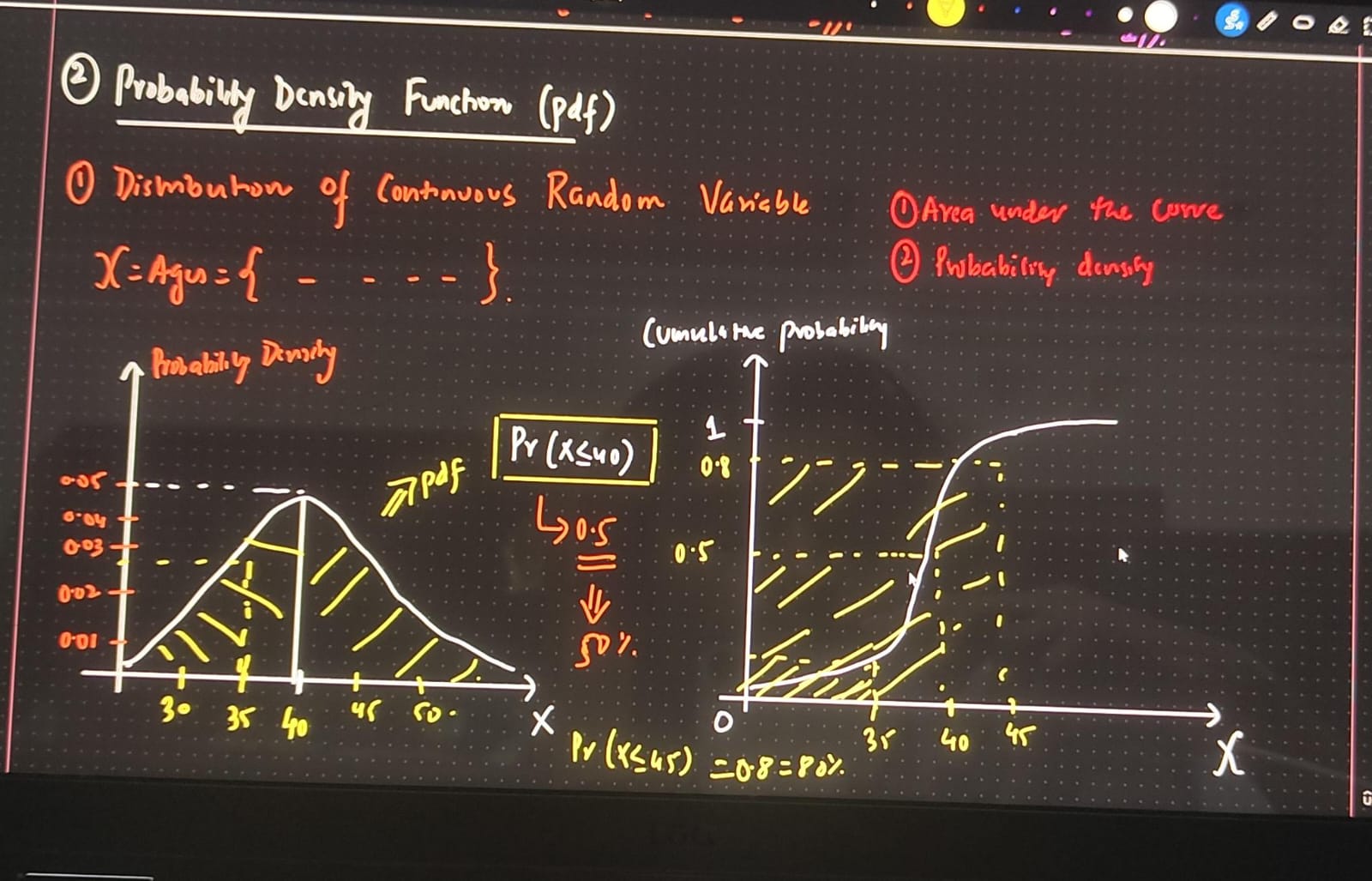
To showcase the distribution of continuous random variable

X = Ages = […]

X axis we have the random variable.

Y axis we have probability density.

We can create a histogram and smoothen it.

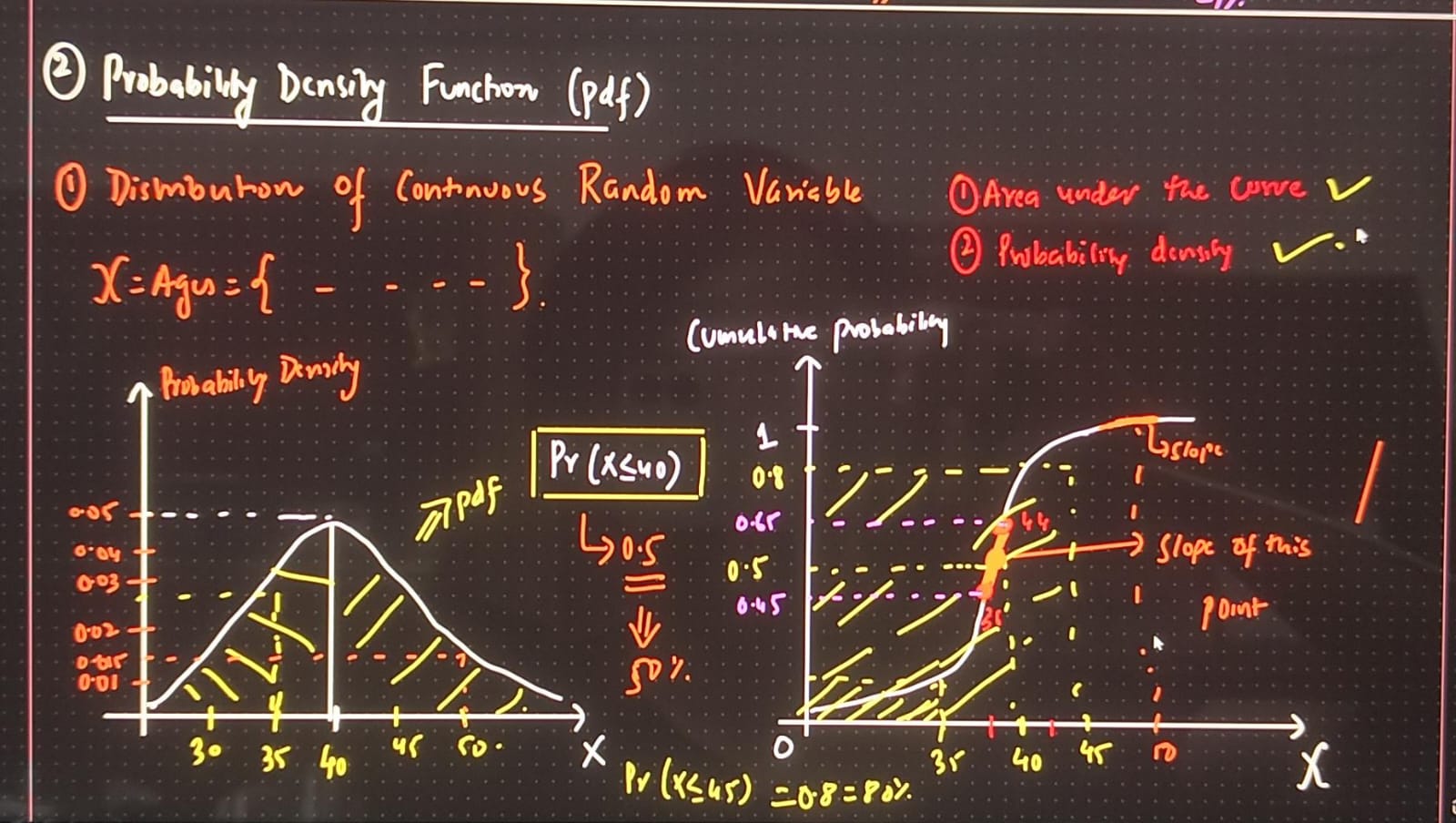


Symmetrical distribution, meaning 50% distribution of the random variable are present on either sides.

40 is the mean.

1. Area under the curve

2. Probability Density



Converting pdf into cds:

On the x axis we have the values of the random variable.

Y axis ranges from 0 to 1.

Cumulative Probability Distribution usually takes an S shape.

According to the cumulative probability distribution

P(X<=40) = 0.5

P(X<=45) = 0.08

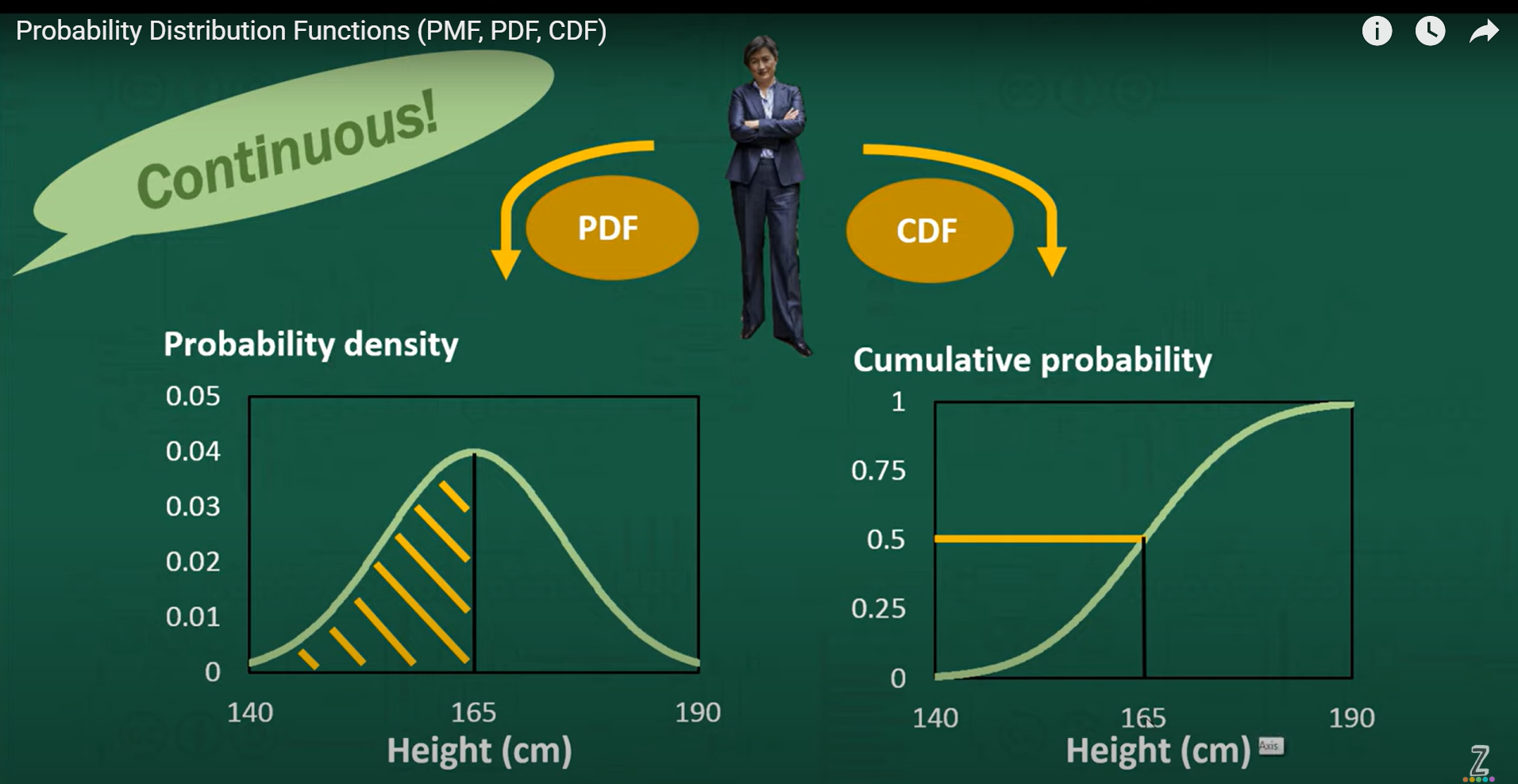
In order to calculate the probability density, we need to find out the gradient or slope of the point.

Steeper the slope, higher the value of the slope.

Slope =

The value calculated, i.e, the slope will be equal to the probability density.

Probability density is calculated by finding out the gradient or the slope.  
Probability density is nothing but gradient of cumulative density function.



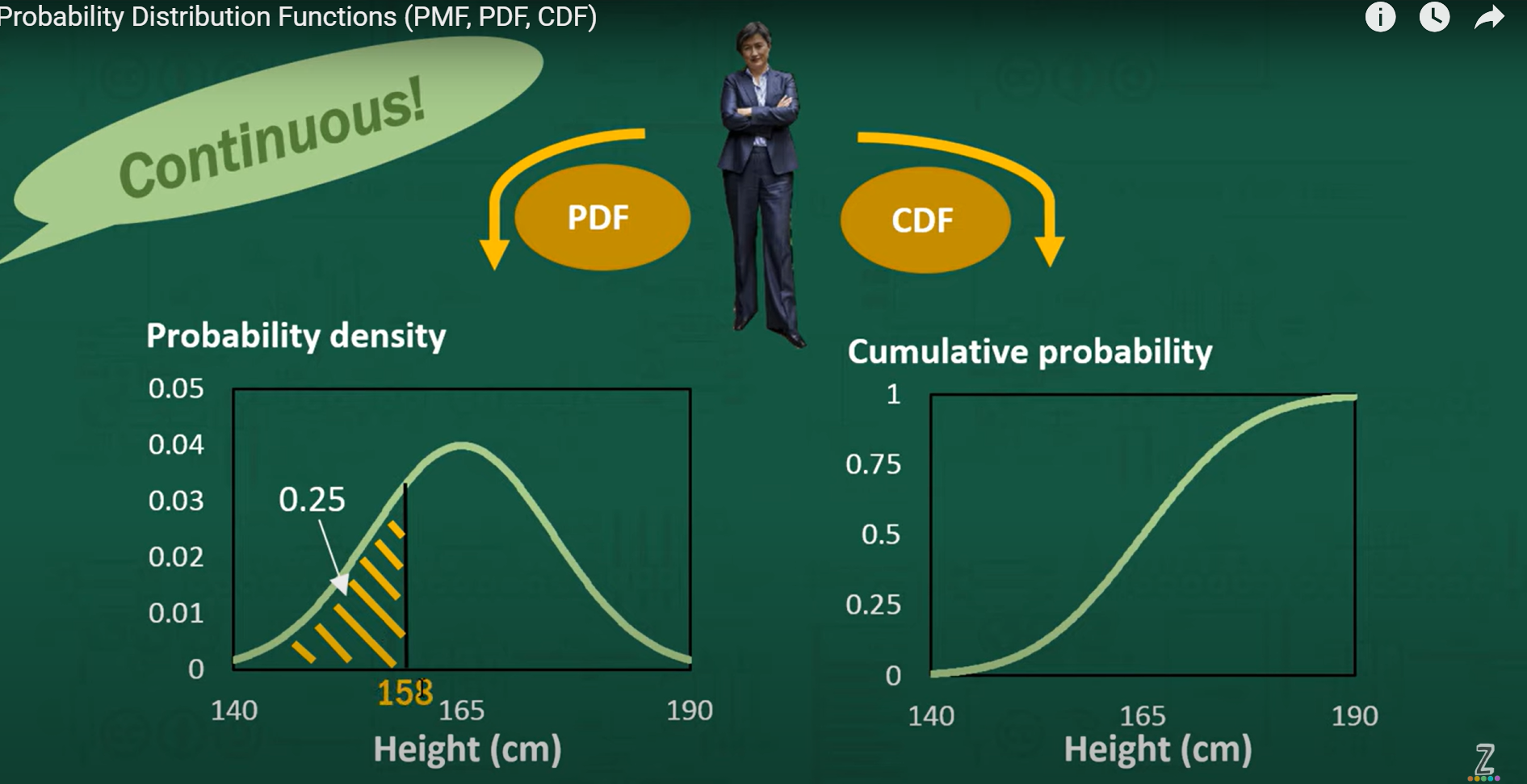
Taking the example of height of women. It is continuous.

Because 165 is mean, it is right in the middle meaning 50% both sides.

If we have a bell curve in pdf, then in cdf it would be S shaped.

It tells here that 50% of the distribution has elapsed at point (value = 165)

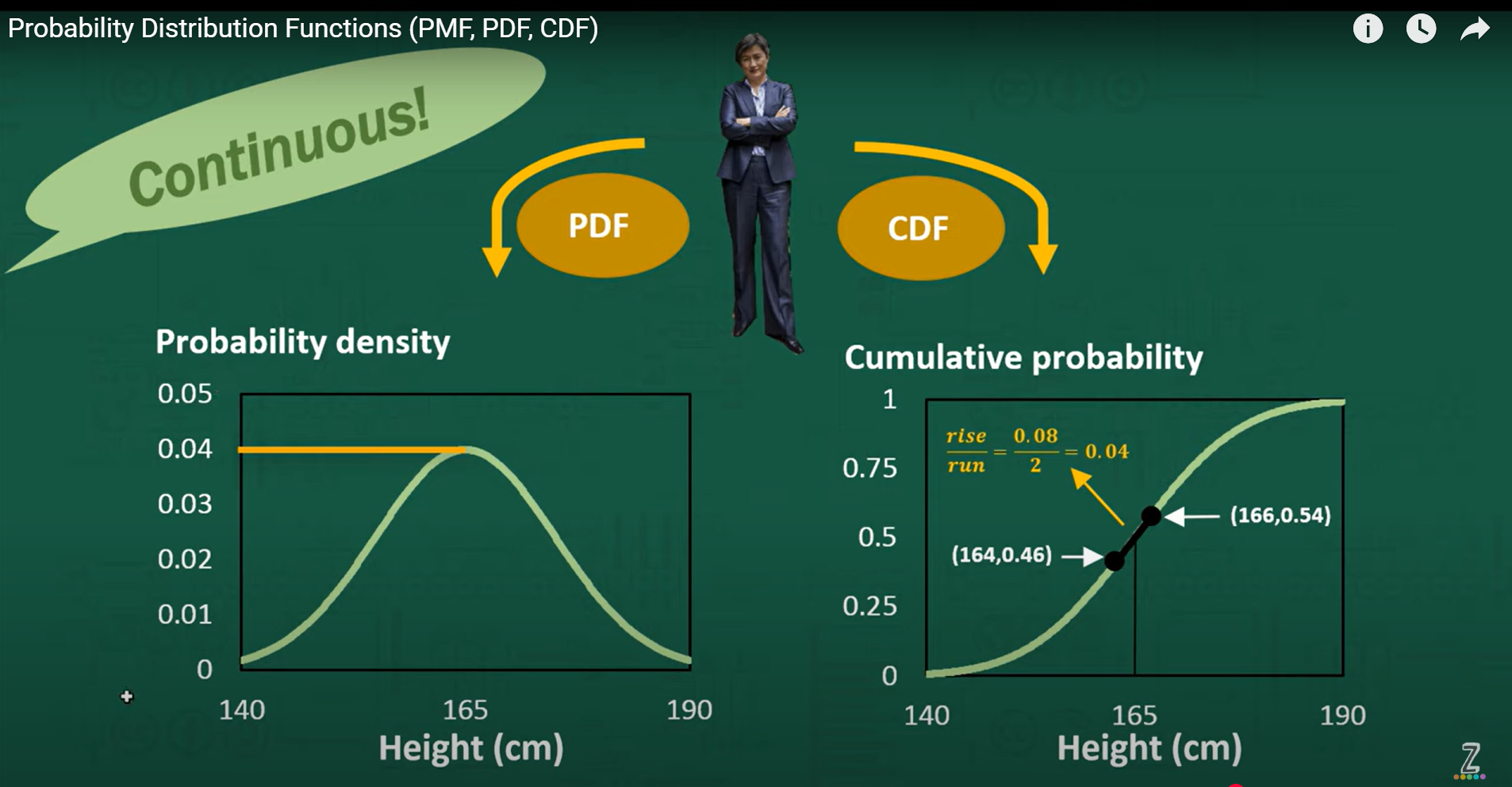
Or we have accumulated half of the distribution by the time we come to height of 165 cm.



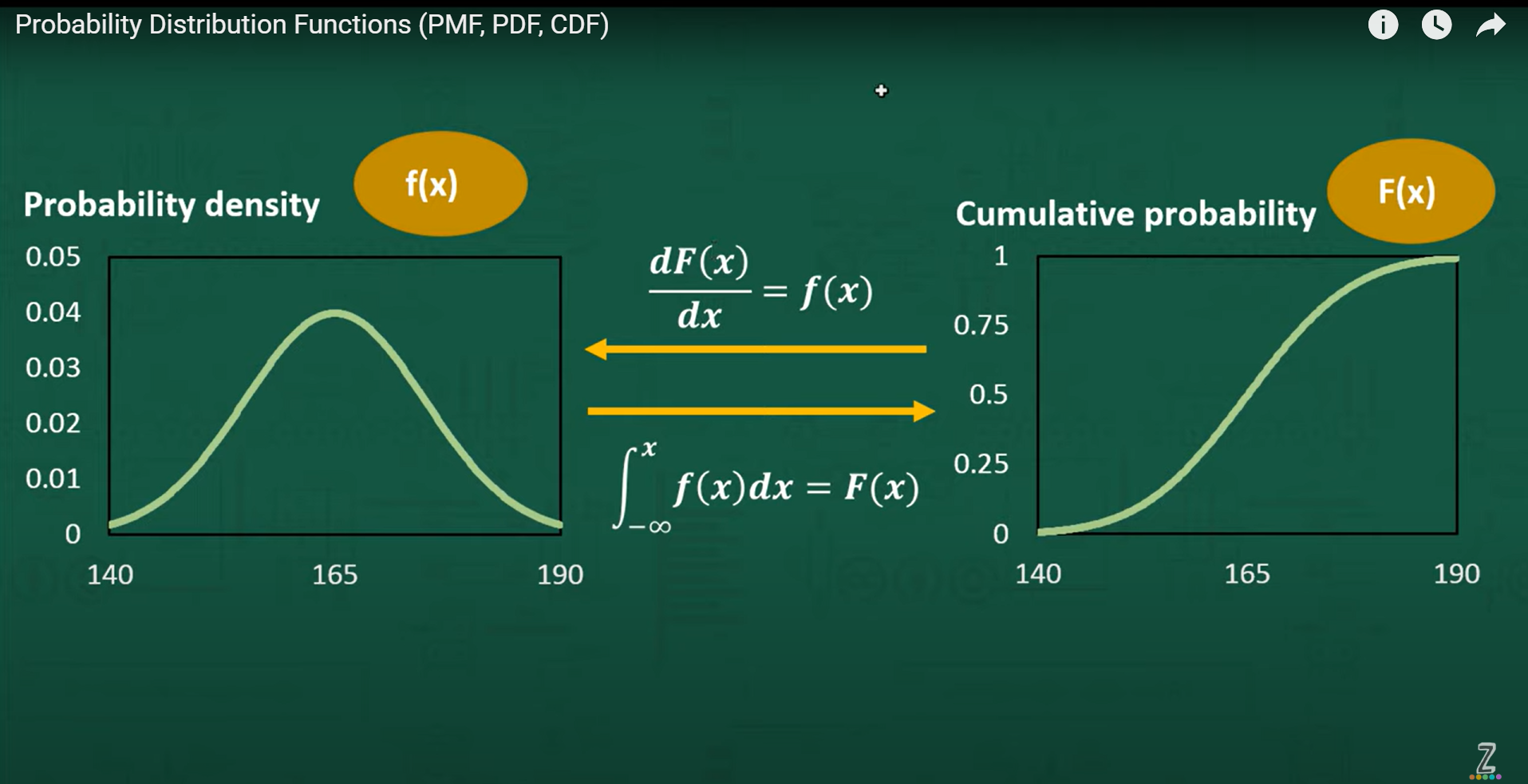
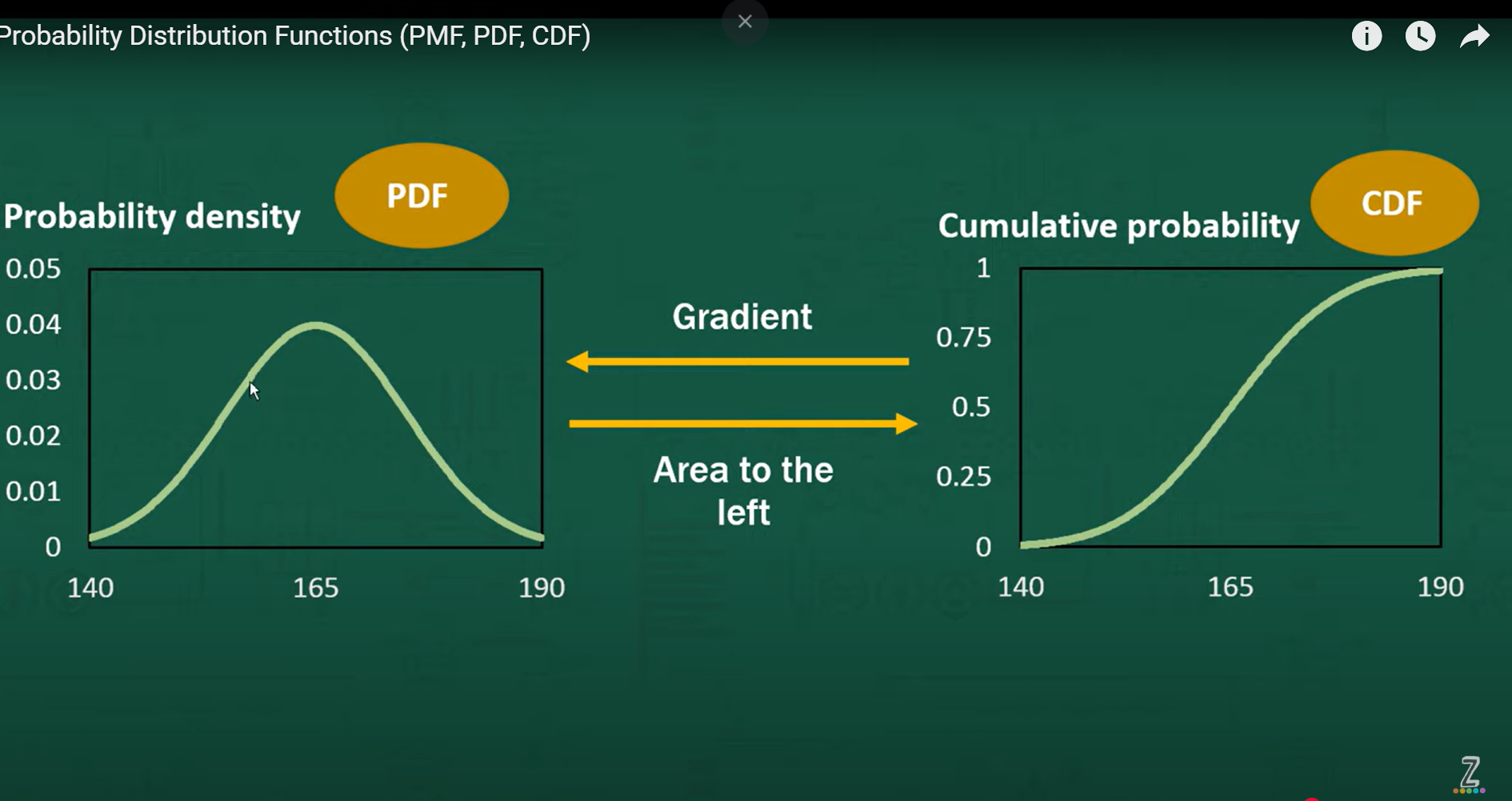
In essence, the values on the y axis, tells us how much of the distribution is to the left of a particular height.

We can create cdf from pdf and pdf from cdf.

Higher gradient = Higher density = More of the distribution is going to be in the area.





   
  
Properties of Probability Density Function:

Non negativity f(x) >= 0 for all x

The total area under the PDF curve is equal to 1

The distribution is created by f(x)

With respect to different distribution, f(x) will change.