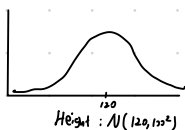
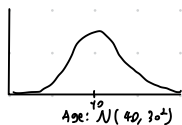


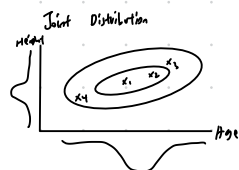
Why do we model data as distribution?

: Imagine you're a criminal, and you want to generate thousands of fake identities. Each fake identity is made up of variables, representing the demographics of a person (Age, Height)



Data	Age	Height
$x_1$	28	198
$x_2$	42	182
$x_3$	43	188
$\vdots$	$\vdots$	$\vdots$
	45	174
	47	182
	42	172
$x_N$	38	168

: it's just one data point (Age, Height)



: To generate fake identities that make sense, you need joint distribution, otherwise you may end up with an unreasonable pair of (Age, Height)

: We can also evaluate probabilities on one of the two variables using conditional probability  $P(x, y)$  and/or by marginalizing  $P(x) = \int P(x, y) dy$

$$P[0 < x < 1 \mid y > 100]$$

Think about MNIST Dataset, where each data point has  $28 \times 28$  dimensions

Then,

data:	$c_1$	$c_2$	$c_3$	$\dots$	$c_{784}$
$x_1$					
$x_2$					
$\vdots$					
$x_N$					

: is one data point.

Hence, MNIST will have 256 independent pdf, and  $P_X$  is a joint distribution of 256 independent pdfs.

