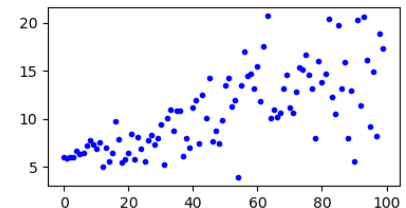
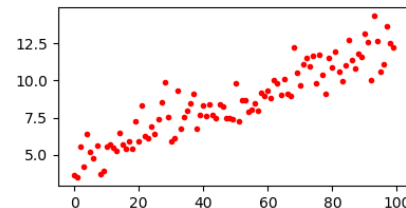
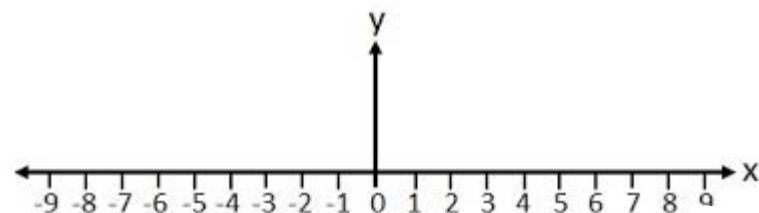


Statistics

Summary statistics

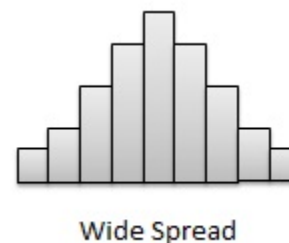
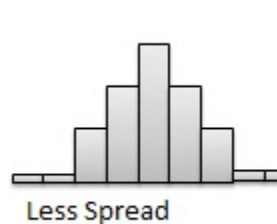
- A dataset is made of variables & observations
- We may want to summarise the distribution of variables for each of the value
- When we explore data we are interested in

Position



Shape

Spread



Summary statistics

Mean / Average

Given sequence:

13, 18, 13, 14, 13, 16, 14, 21, 13

The mean is the usual average, so:

$$(13+18+13+14+13+16+14+21+13) / 9 = 15$$

Median

Order the list : 13, 13, 13, 13, 14, 14, 16, 18, 21

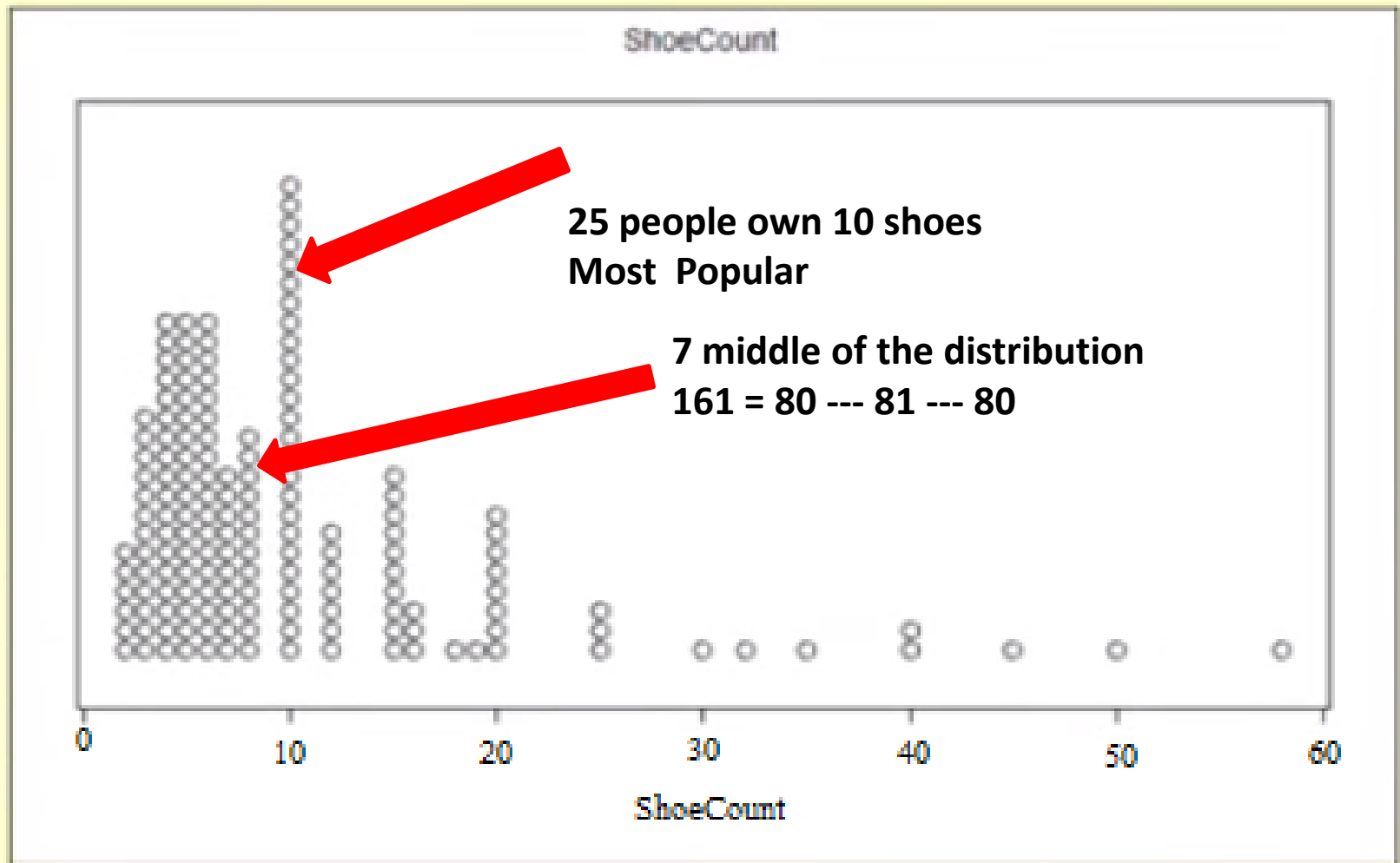
There are **nine numbers** in the list, so the middle one will be $(9+1)/2 = 10/2 = 5$

= 5th number, So the median is 14.

Mode

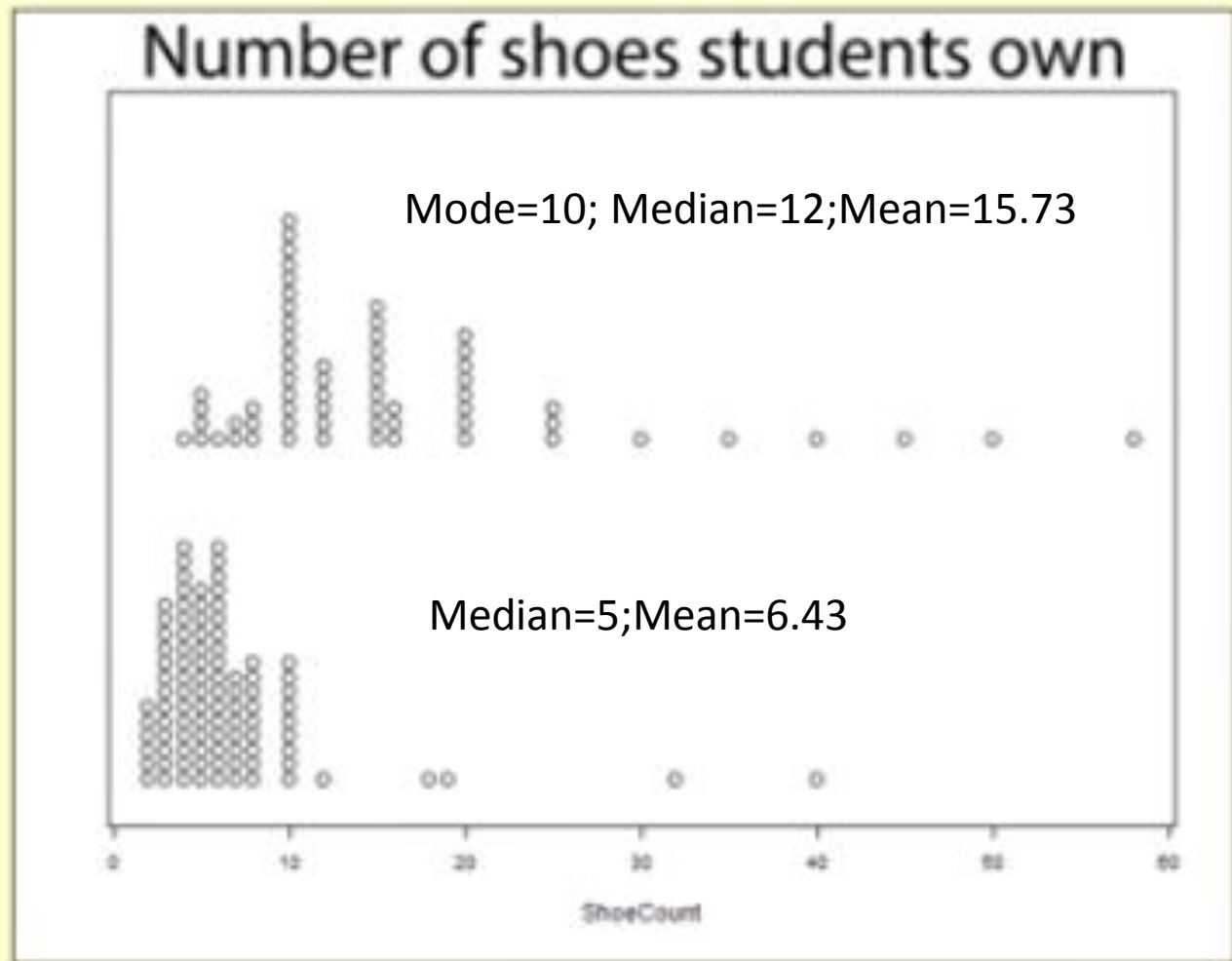
The mode is the number that is repeated more often than any other, so 13 is the mode.

Mode=10;Median=7;Mean=10.07



Female

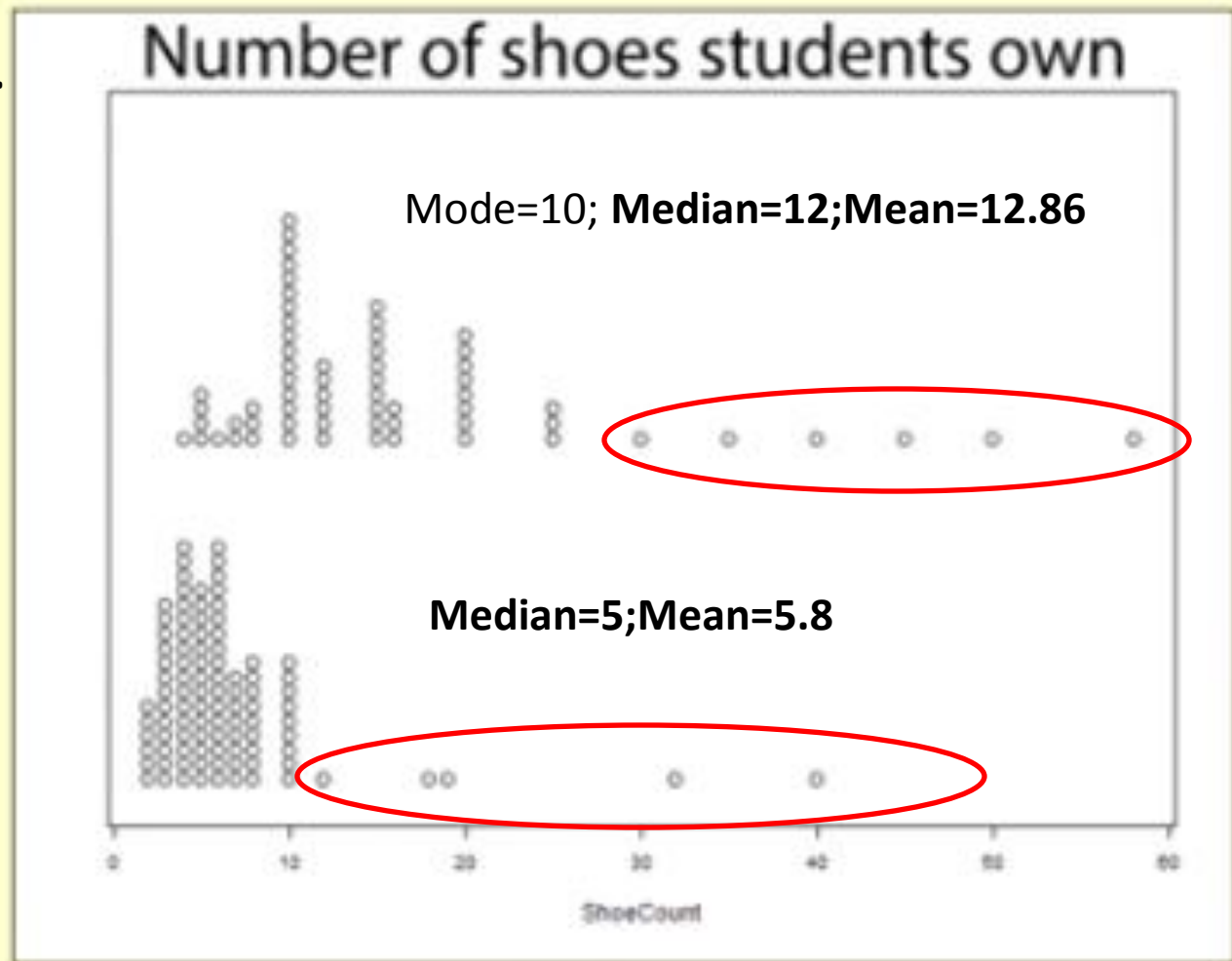
Male



Both cases median remains unchanged.
For this example
MEDIAN is better
indicative

Female

Male



Statistical Analysis

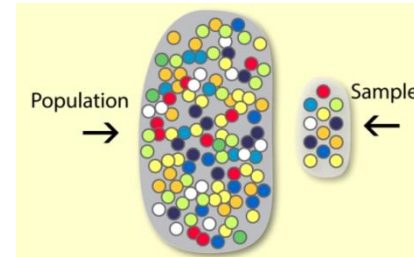
Inferential statistics

The process of **drawing conclusions** about **population** parameters based on a **sample** taken from the **population**

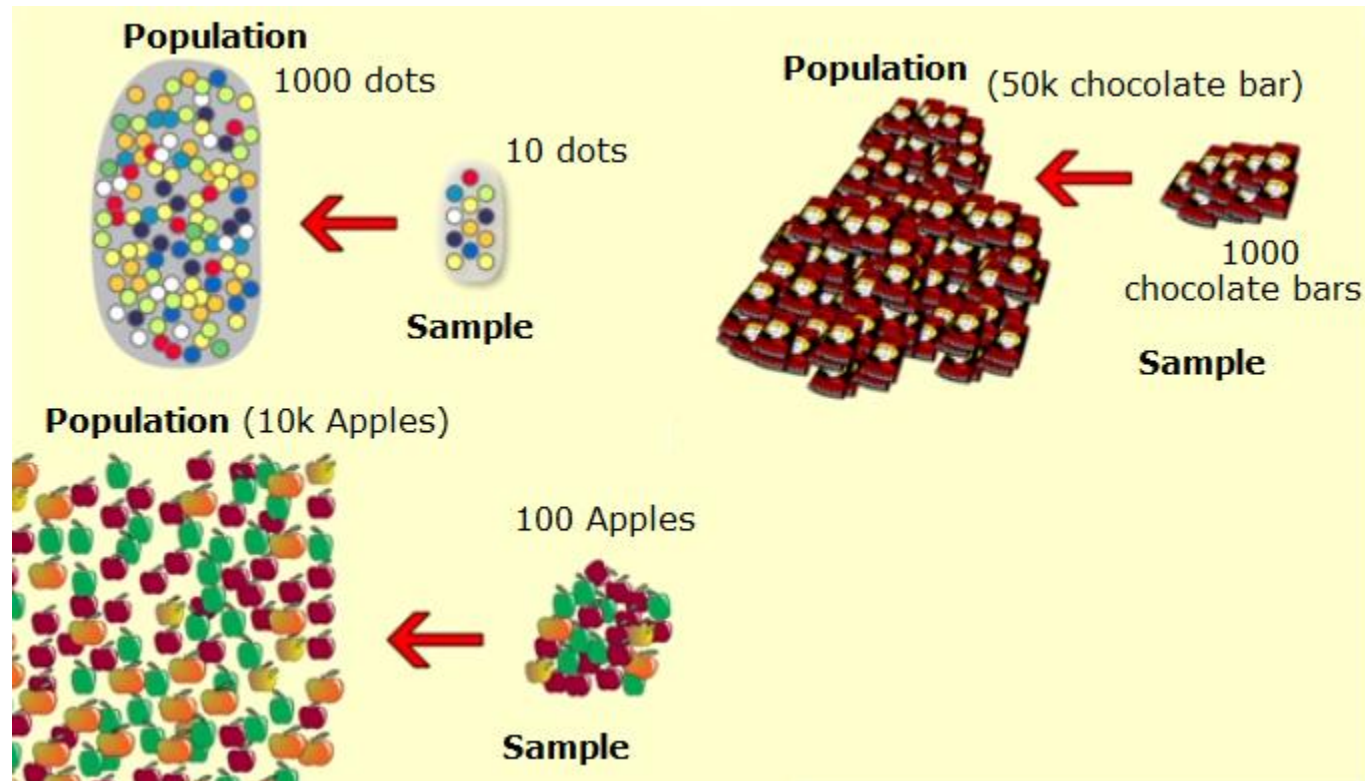
Allows us to make conclusion beyond the data we have to the population to what it was drawn

Samples

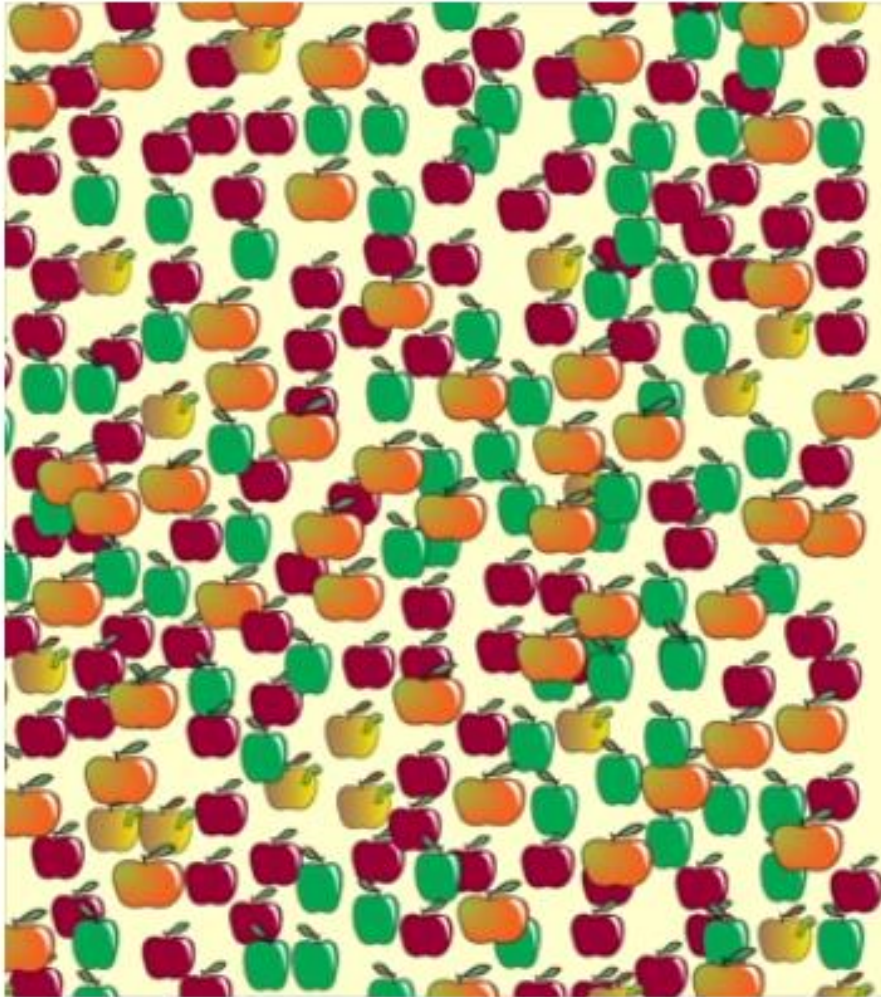
Sample is the data that we collect from the population



Samples help us draw some conclusions about the objects in the population



Samples



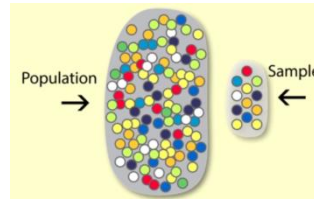
Sample mean = 149 g



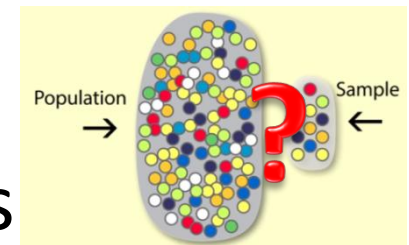
Population mean
is about 149 g

Ideas underling Inference

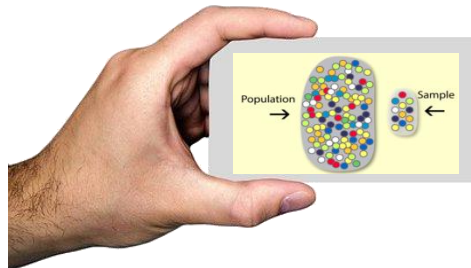
1. A sample is likely to be good representation of the population



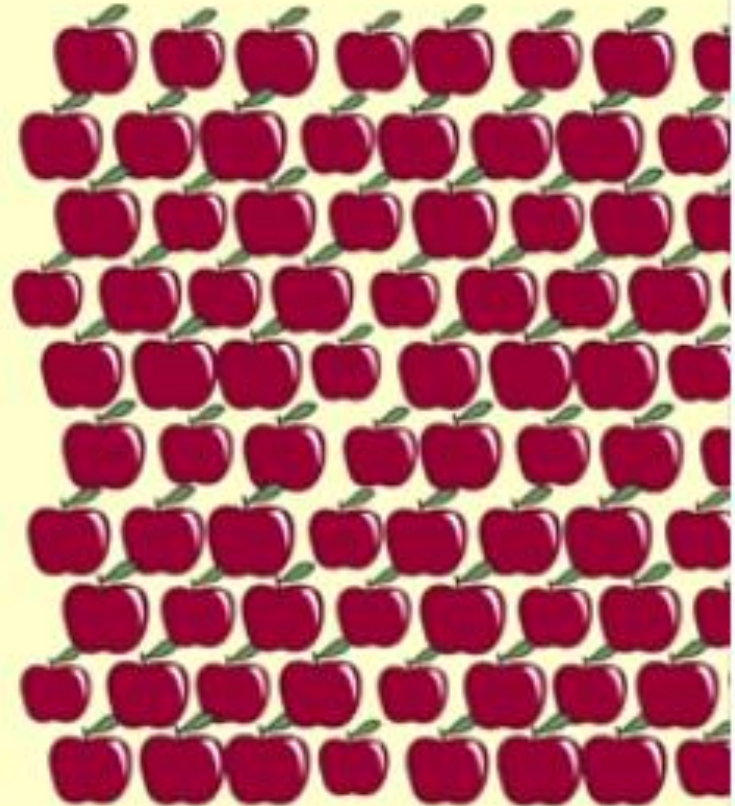
2. There is a element of uncertainty as to how well the sample represents the population



3. The way the sample is taken matters



1. A sample is likely to be good representation of the population



2. There is a element of uncertainty as to how well the sample represents the population



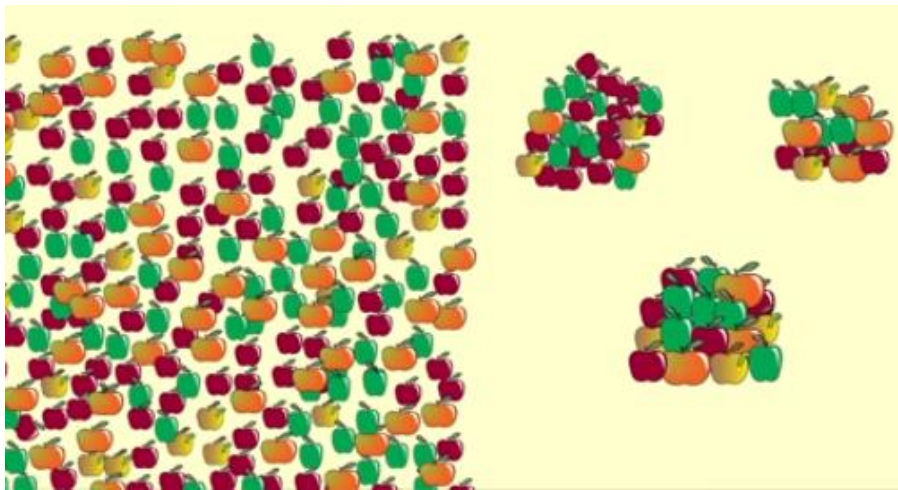
Population



Sample

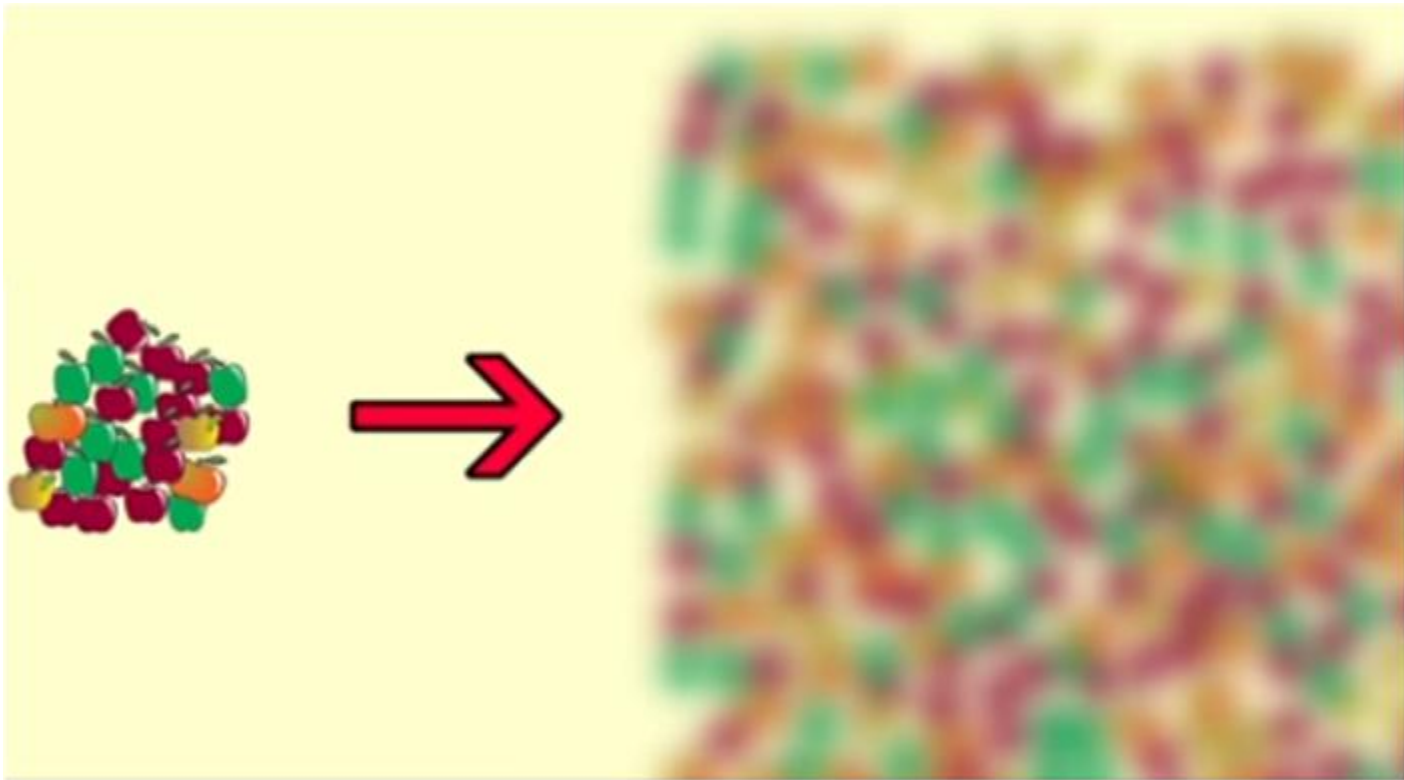
Sample may not always be a perfect representation of what it was drawn from.

Sampling Error



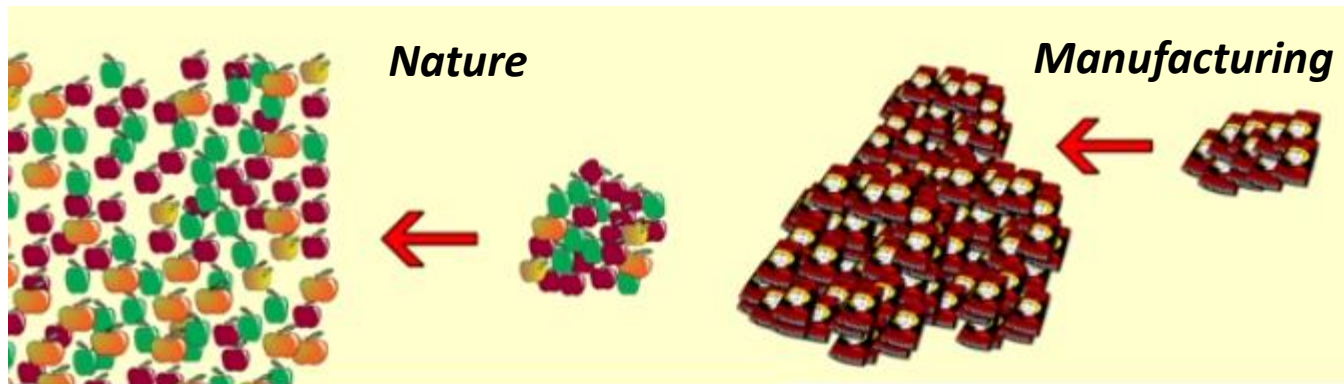
All the samples drawn from the same population may not be the same

From Simulation and probability theory we can get an idea of what the population is likely to be like from the information the sample gives us



3.The way the sample is taken matters

The sample must be representative of the population, this happens when each person in the population has a equal chance of being selected in the sample

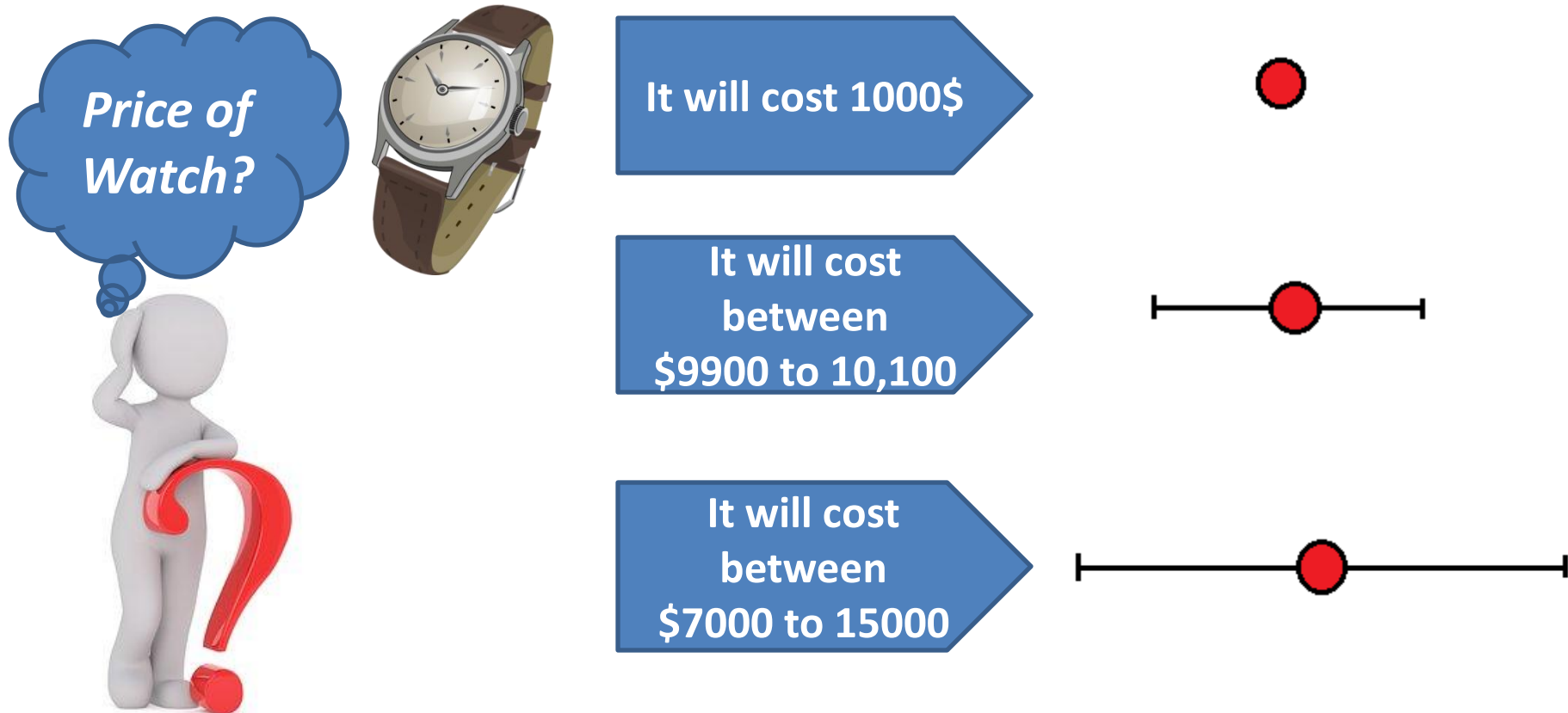


Difficult to Sample

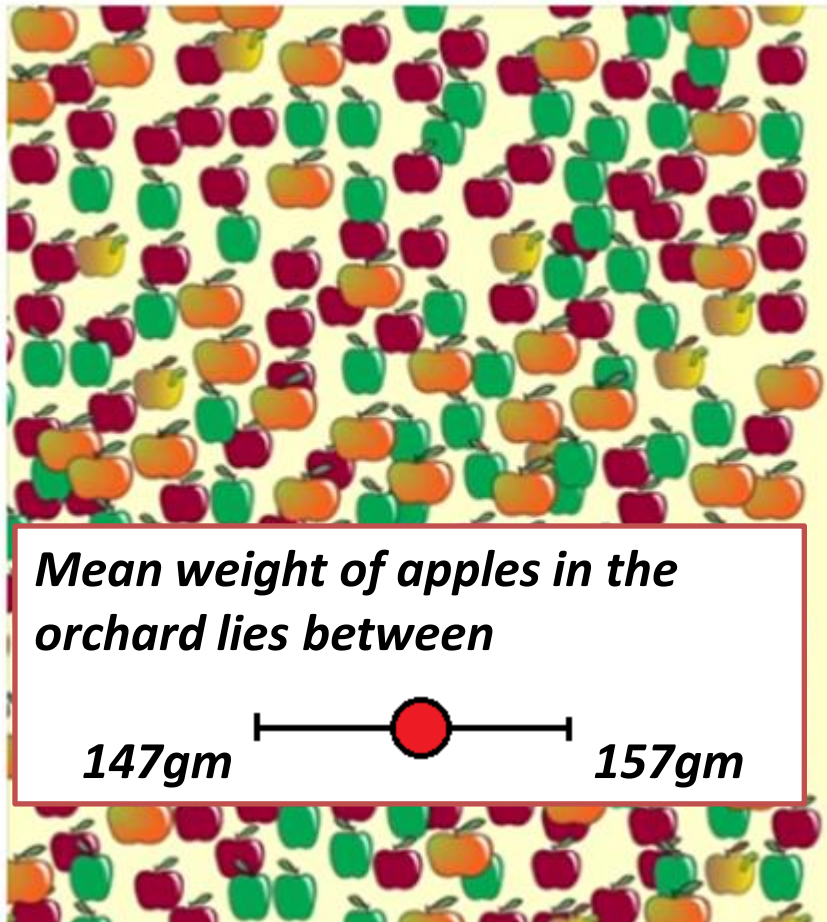
Confidence Interval

Confidence Interval indicates how accurate our estimate lies to be

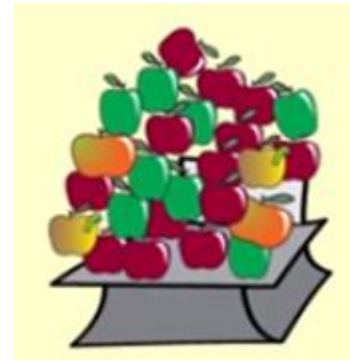
Confidence Interval



What is the mean weight of Apples in the Orchard?



Take Sample and Calculate mean



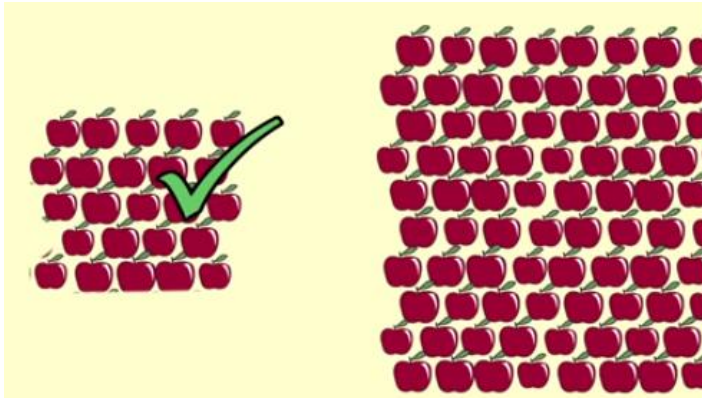
Sample mean = 149 g



Population mean
is about 149 g

What affects the width of the CI

1. Variation within the population of interest



2. Size of the sample



139gm



182gm



280gm



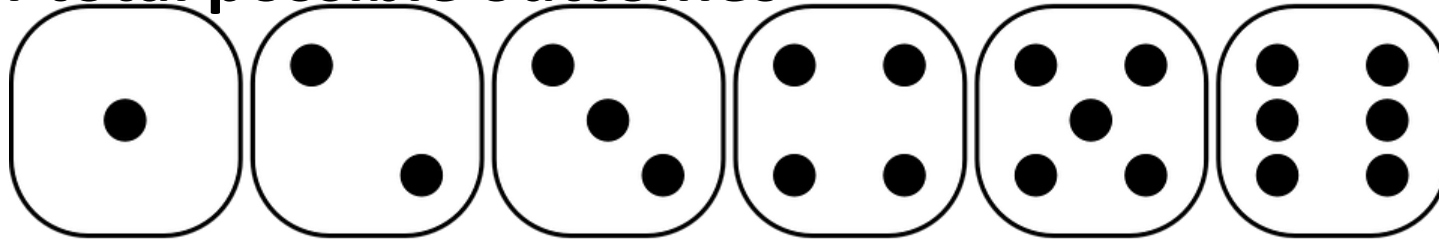
310gm



Probability

of outcomes you looking for

of total possible outcomes



Probability of rolling a dice

1 outcome / 6 total possible outcome

= 1/6

Probability of rolling even number

3 outcome / 6 total possible outcome

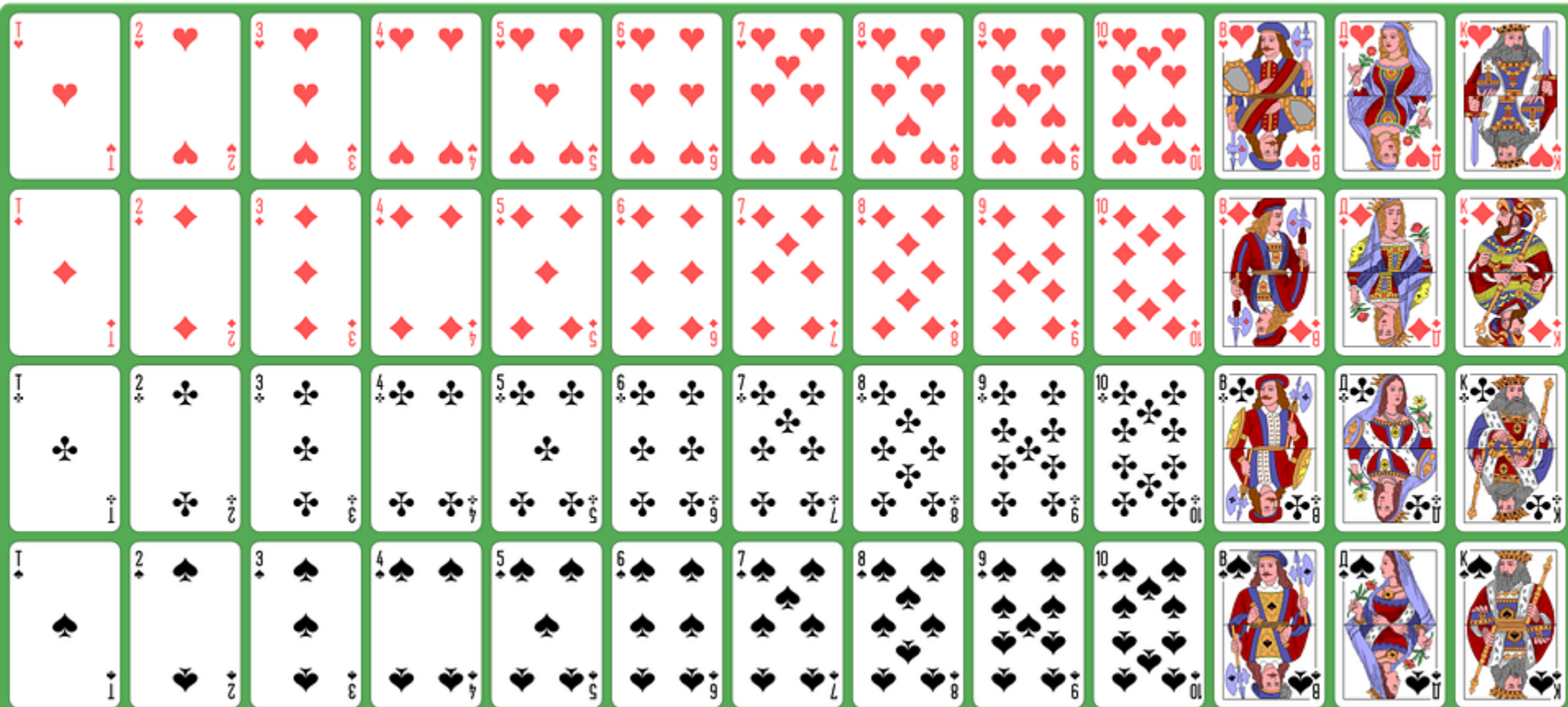
= 3/6 = 1/2

*Probability of randomly drawing a heart
= $13/52 = 1/4$*

Probability of drawing queen = $4/52$

Probability of drawing queen of hearts = $1/52$

Probability of drawing face card = $12/52$



Conditional Probability

Events

A: Draw a Heart

B: Draw a face card

$$P(A) = 13/52$$

$$P(B) = 12/52$$

*Draw a Face card
that's a Heart*

$$P(B|A) = 3/13$$

Probability of B given A

