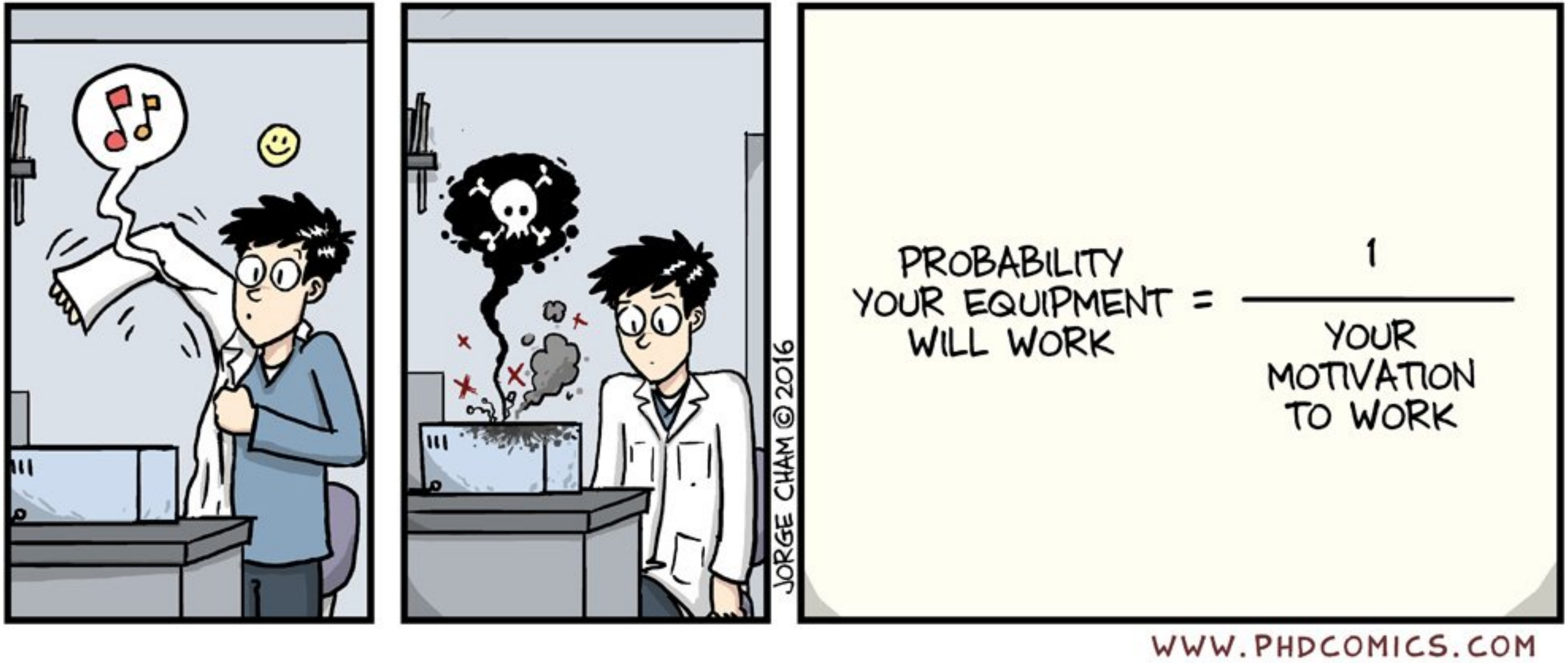


Lecture 02

9.23.21

Refresher Quiz

Describe the **shape**, **center**, and **spread** of the following data.
Sketch a quick graph.

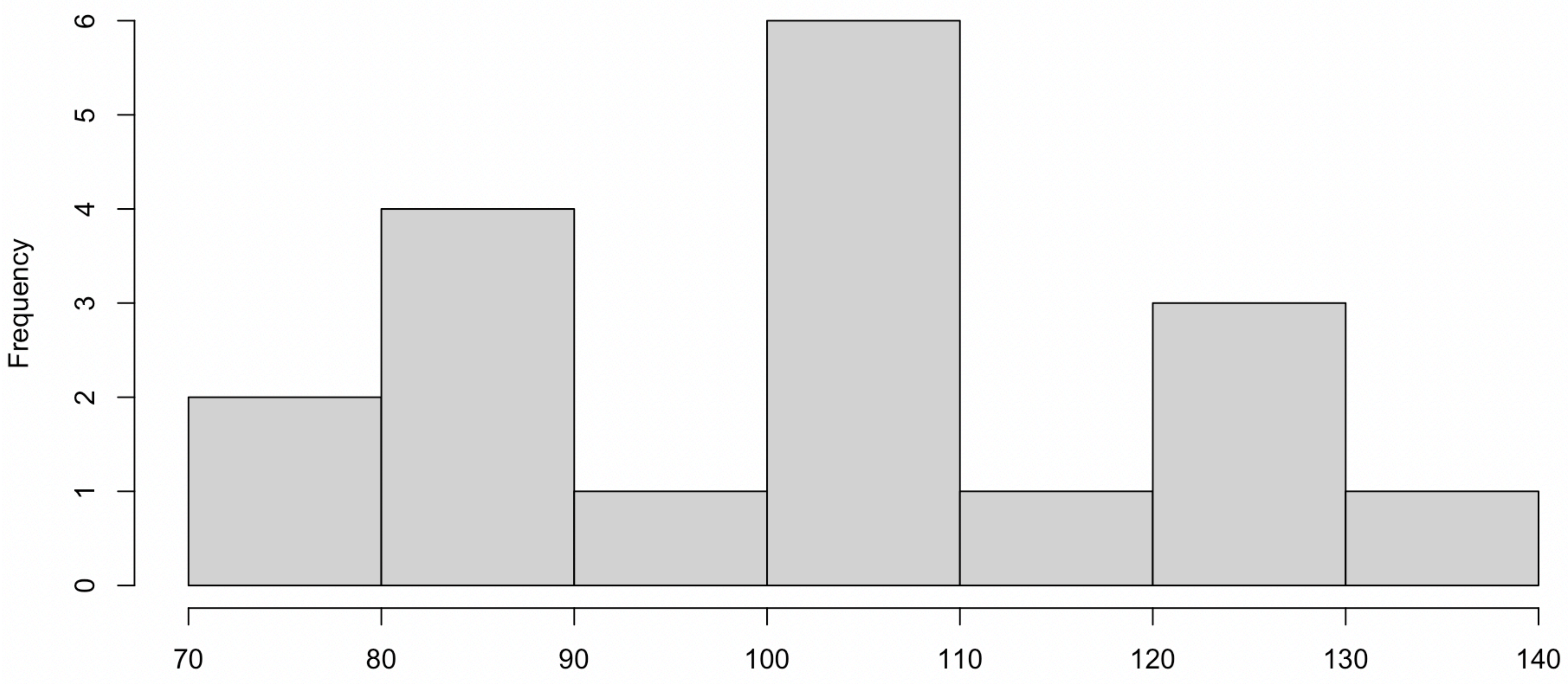
Concentration of calcium (nM) in blood samples from 18 healthy individuals:

102	112	115	107	78	122	110	130	120
112	88	112	103	122	121	125	80	86

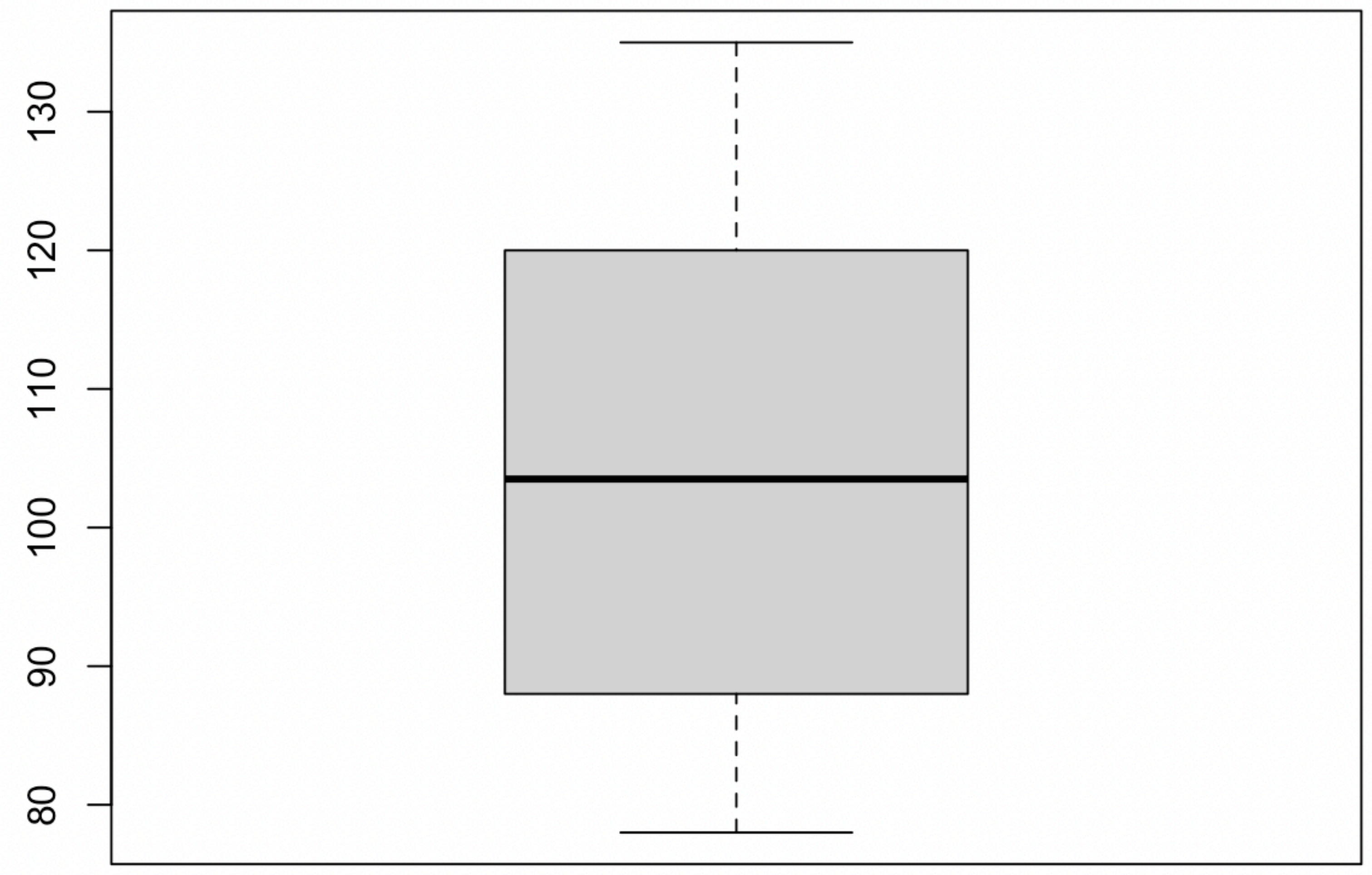
Describe the **shape**, **center**, and **spread** of the following data.
Sketch a quick graph.

Concentration of calcium (nM) in blood samples from 18 healthy individuals:

102	112	115	107	78	122	110	130	120
112	88	112	103	122	121	125	80	86




```
> hist(data)
```




```
> boxplot(data)
```

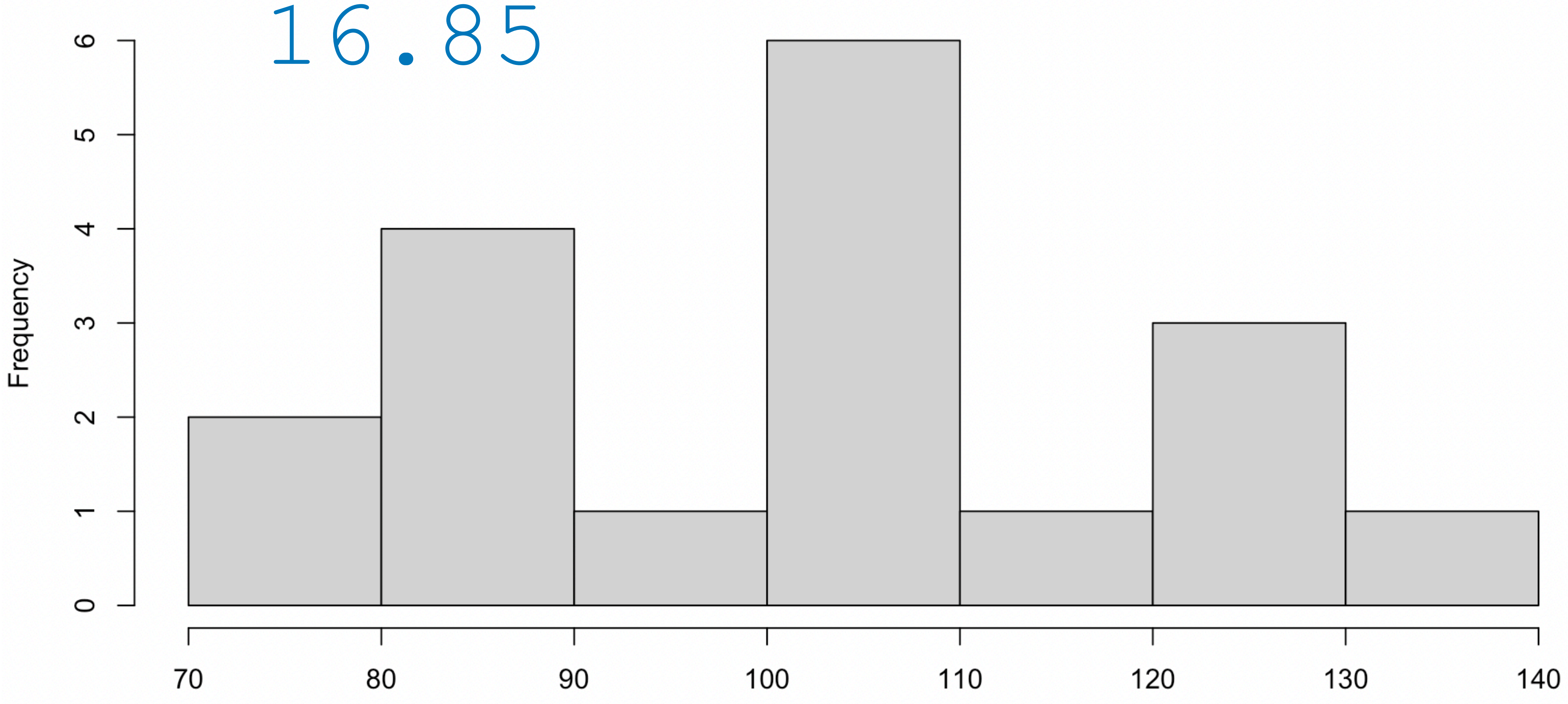
Describe the **shape**, **center**, and **spread** of the following data.
Sketch a quick graph.

```
> summary(data)
```

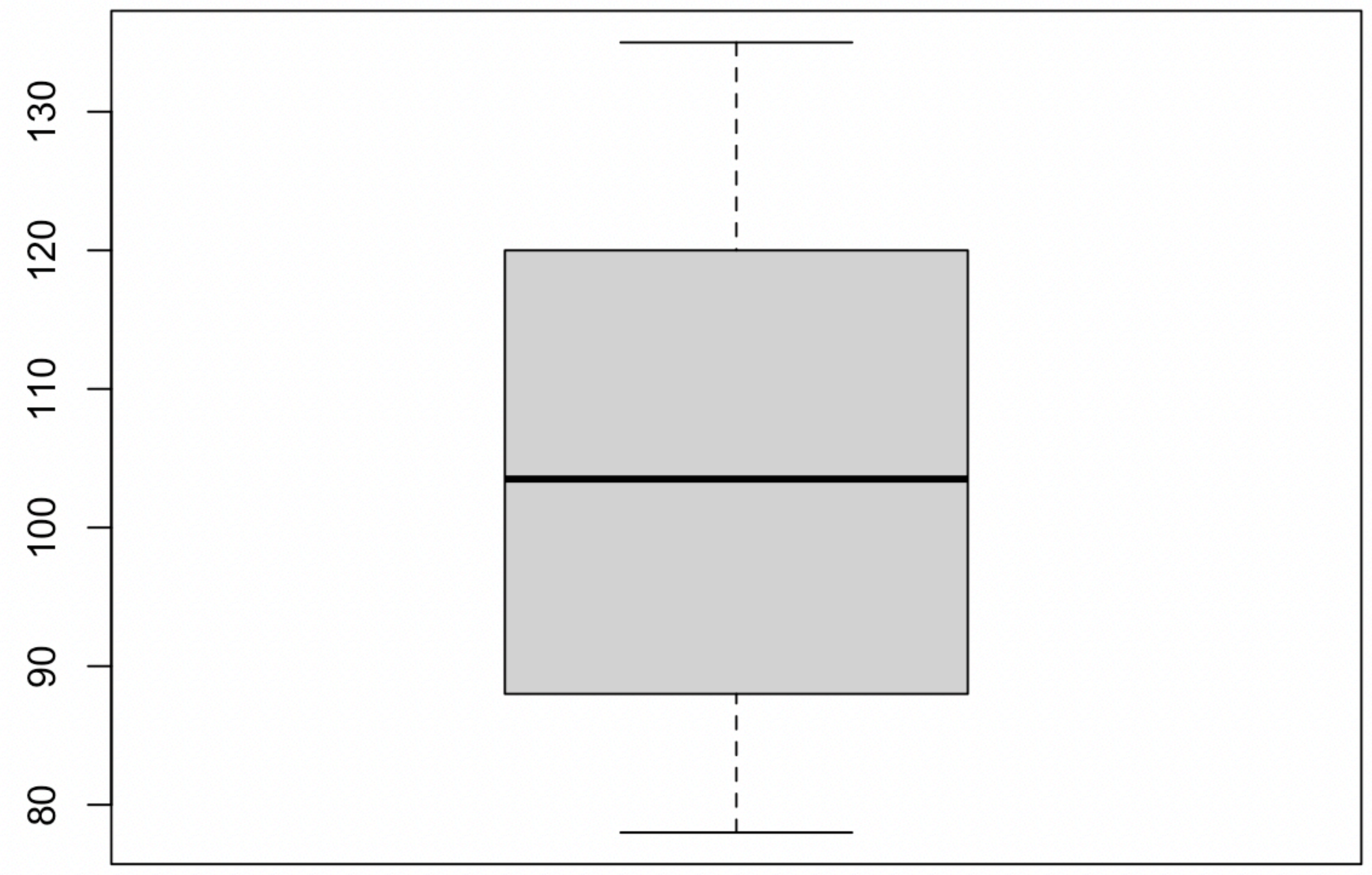
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
78.0	88.5	103.5	103.6	116.8	135.0

```
> sd(data)
16.85
```

IQR? Range? Var?

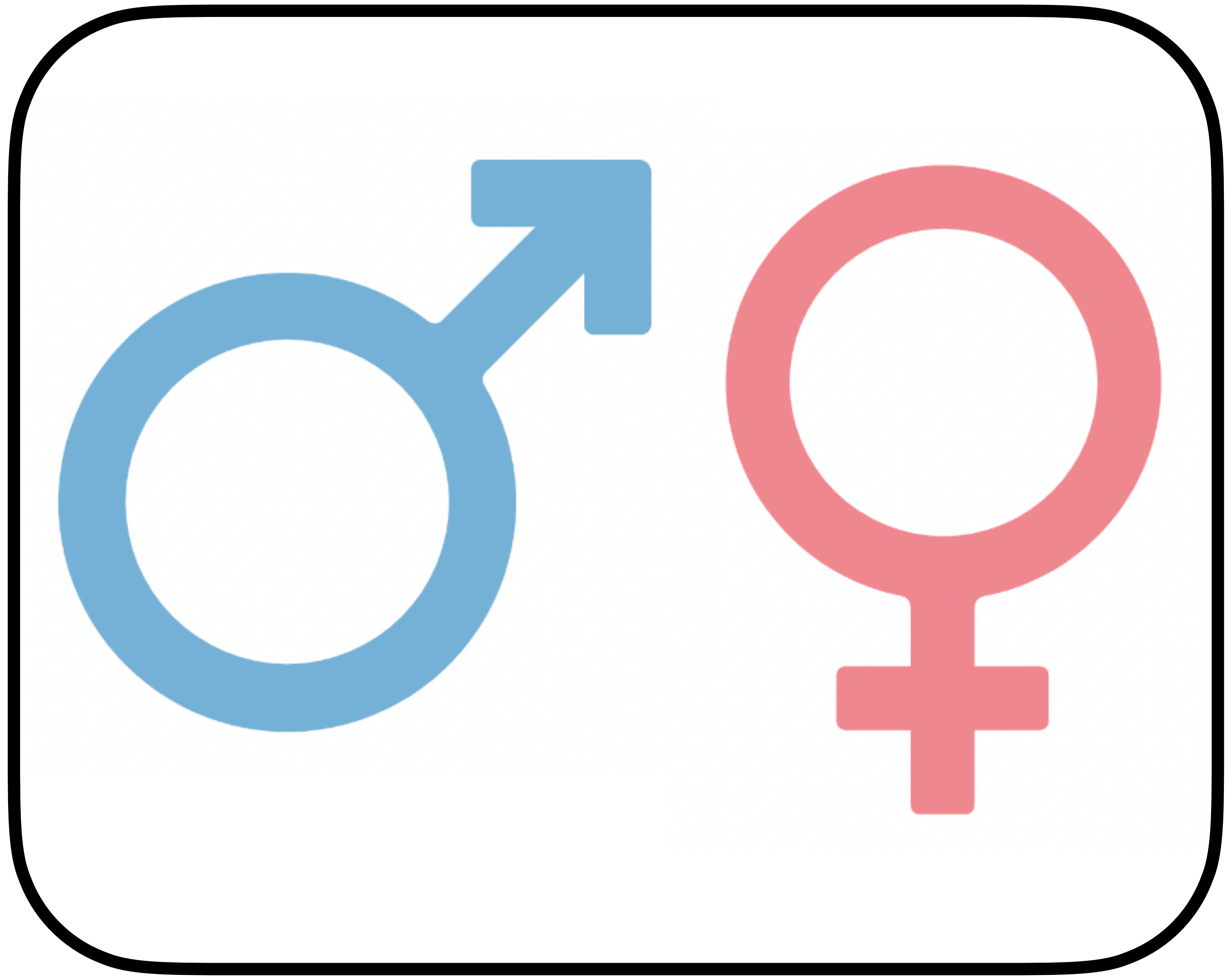



```
> hist(data)
```

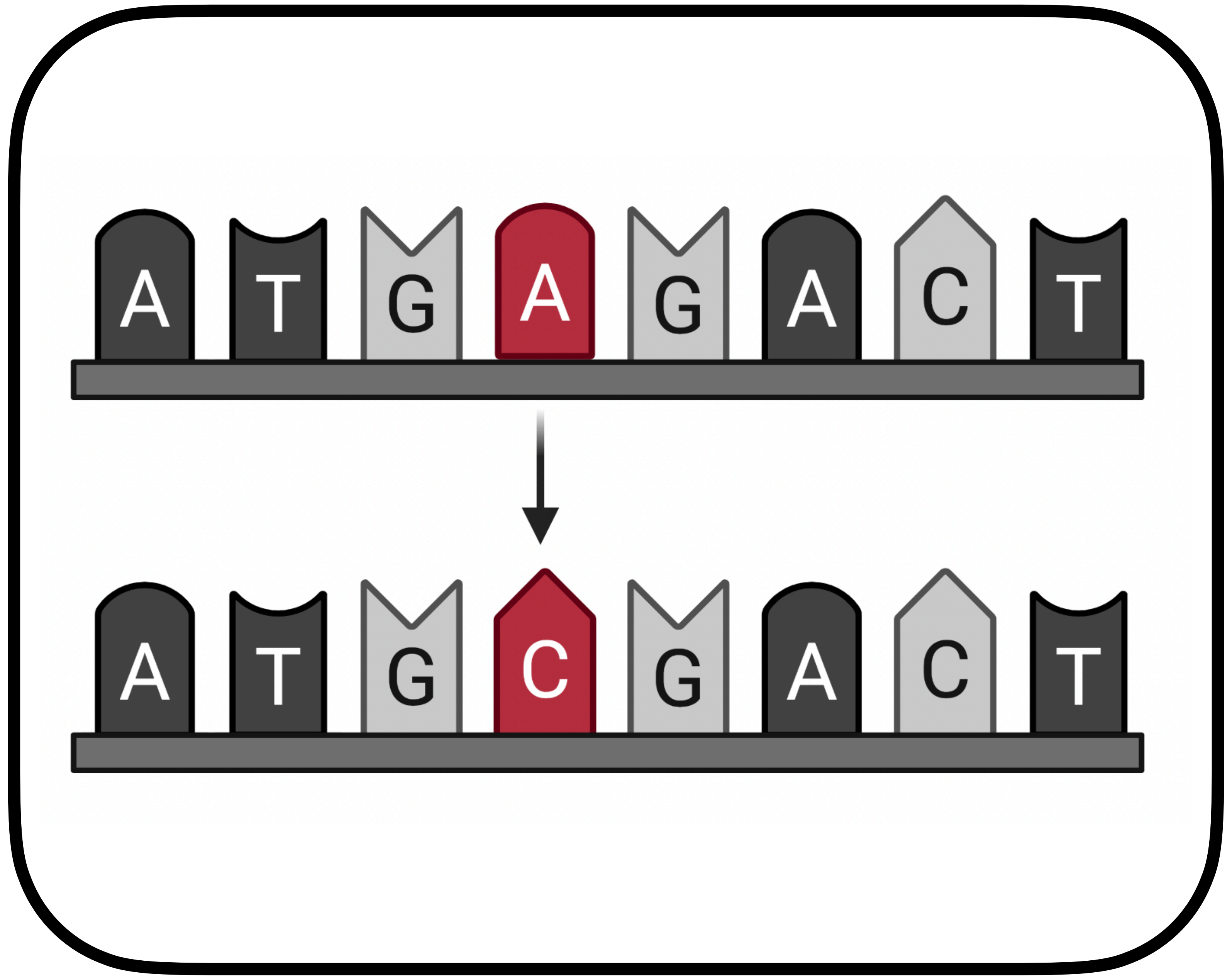



```
> boxplot(data)
```

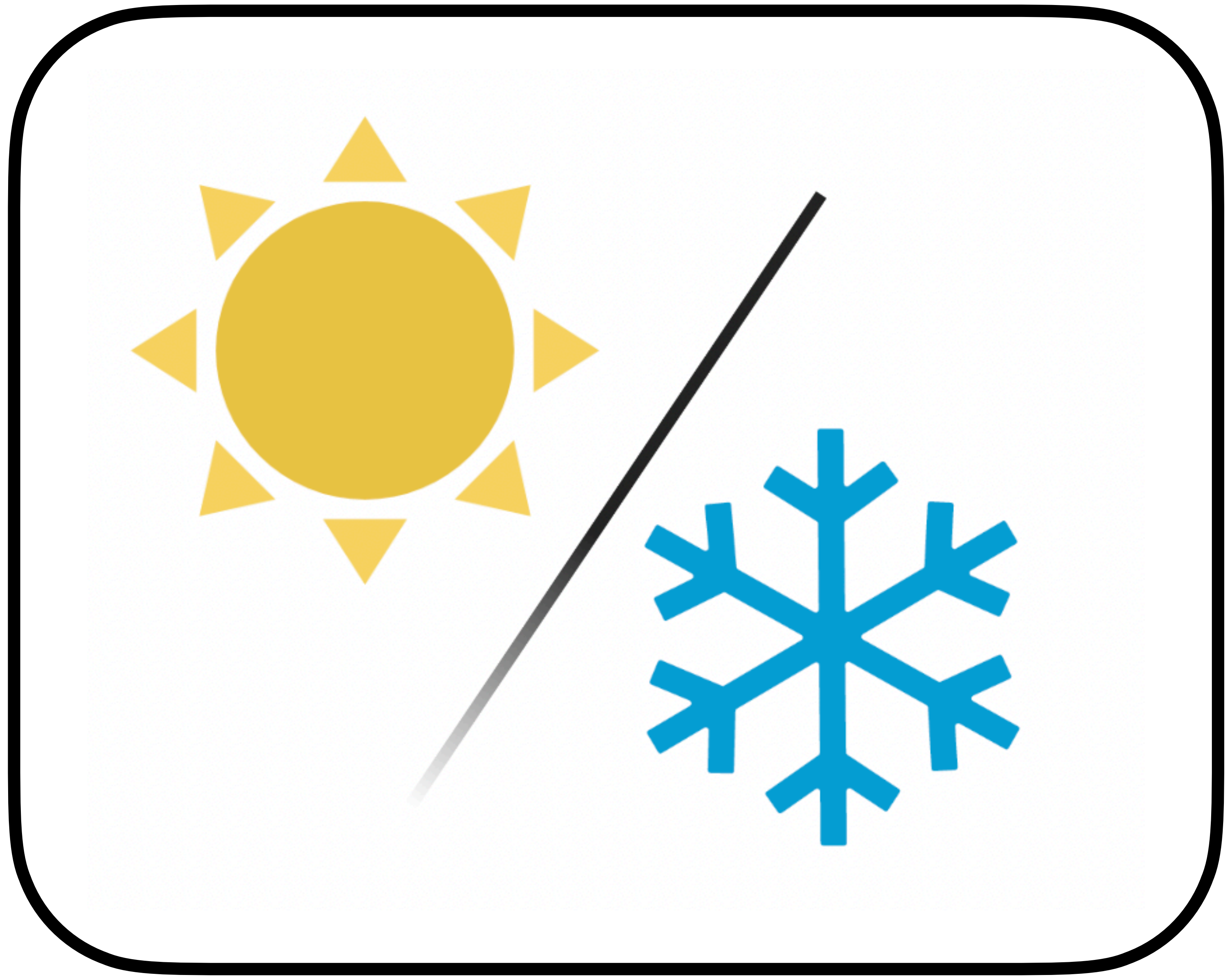
“Chance” is a normal part of biology



Independent assortment

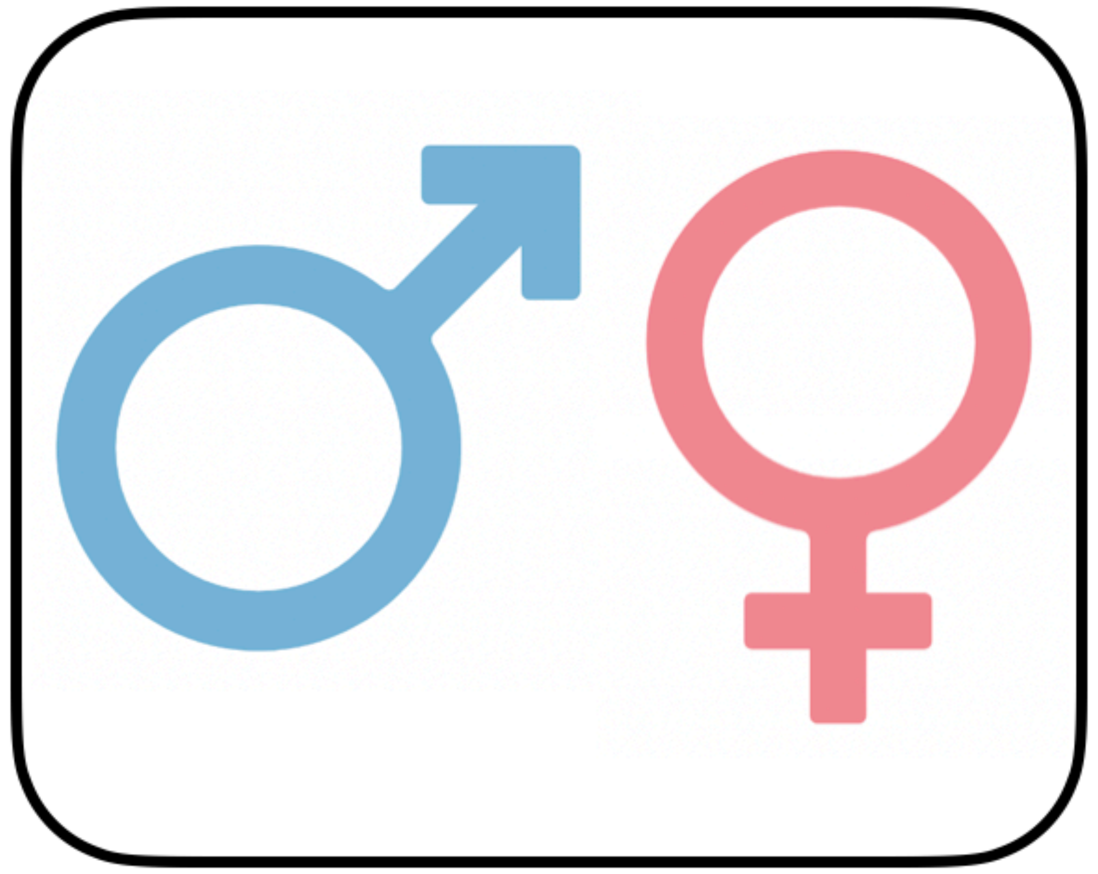


Occurrence of mutation

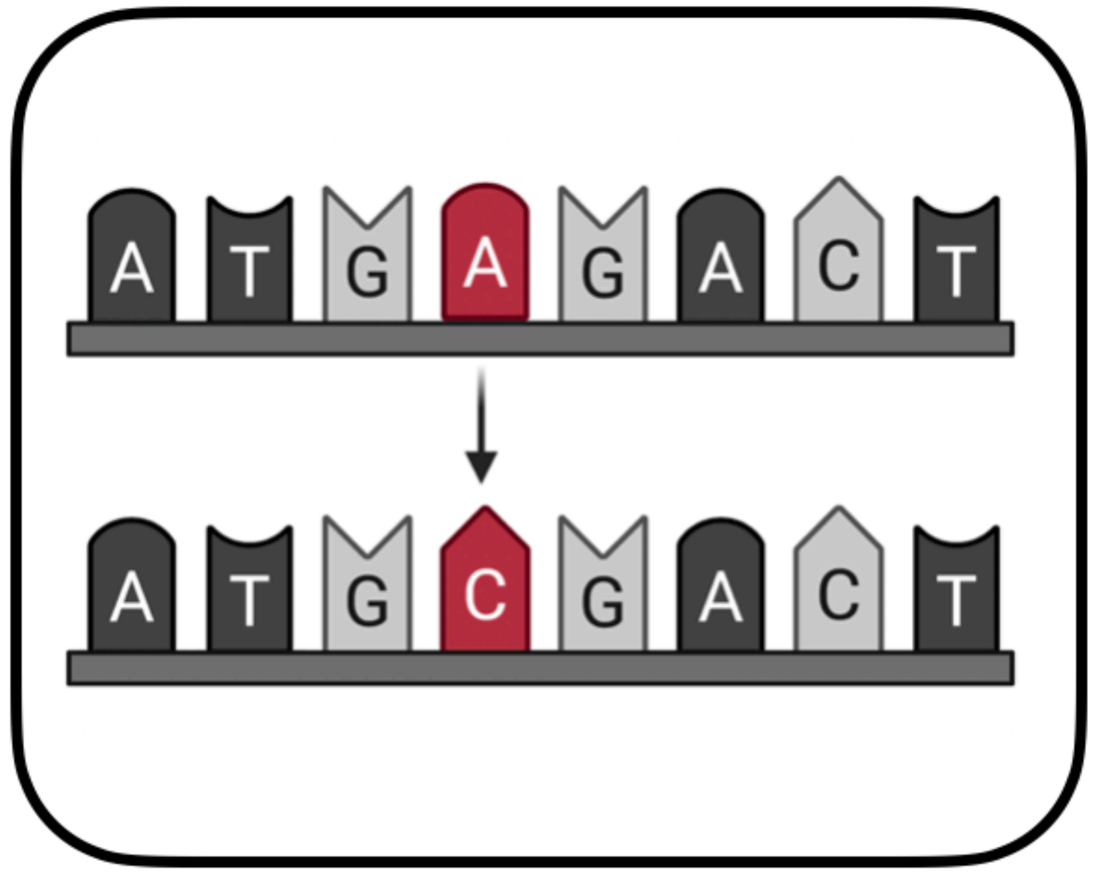


Environmental variation

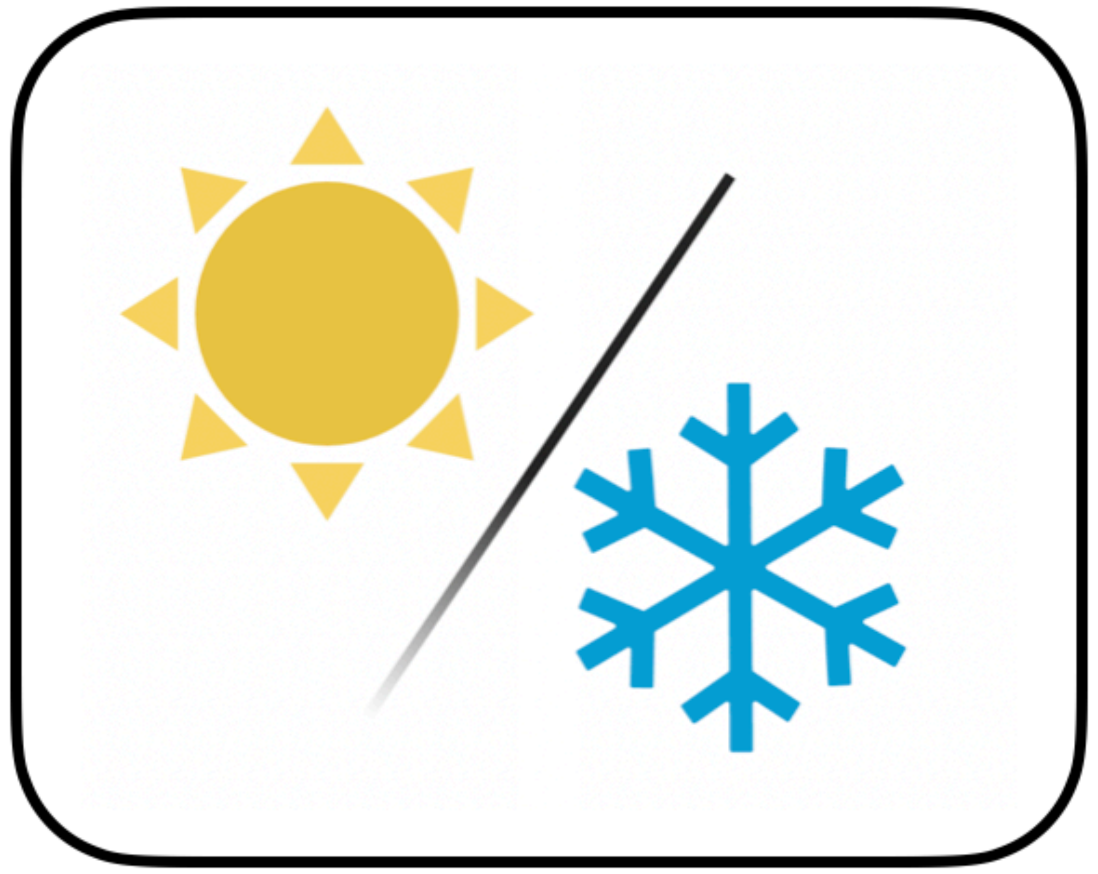
“Chance” is a normal part of biology



Independent assortment



Occurrence of mutation



Environmental variation

Probability

Allows us to quantify how likely (or unlikely) a result is

Probability shows the likelihood of an event

$\boxed{\Pr\{E\} = \text{probability of event } E}$

Number between 0 and 1

E = chance event,
either occurs or doesn't
occur

Example: Probability of tossing a coin

HEADS

TAILS

- **Event:** tossing a coin
- **Outcomes:** 2 options (heads or tails)

Example: Probability of tossing a coin

Pr{HEADS} = ?

$$= \frac{\text{\# Fav. outcomes}}{\text{\# Total outcomes}}$$

$$= \frac{1}{2}$$

HEADS

TAILS

- **Event**: tossing a coin
- **Outcomes**: 2 options (heads or not heads)

Probability = relative frequency (large sample)

Pr{HEADS} = ?

$$= \frac{\textbf{\# HEADS}}{\textbf{\# Total tosses}}$$

HEADS

TAILS

- **Event**: tossing a coin
- **Outcomes**: 2 options (heads or not heads)

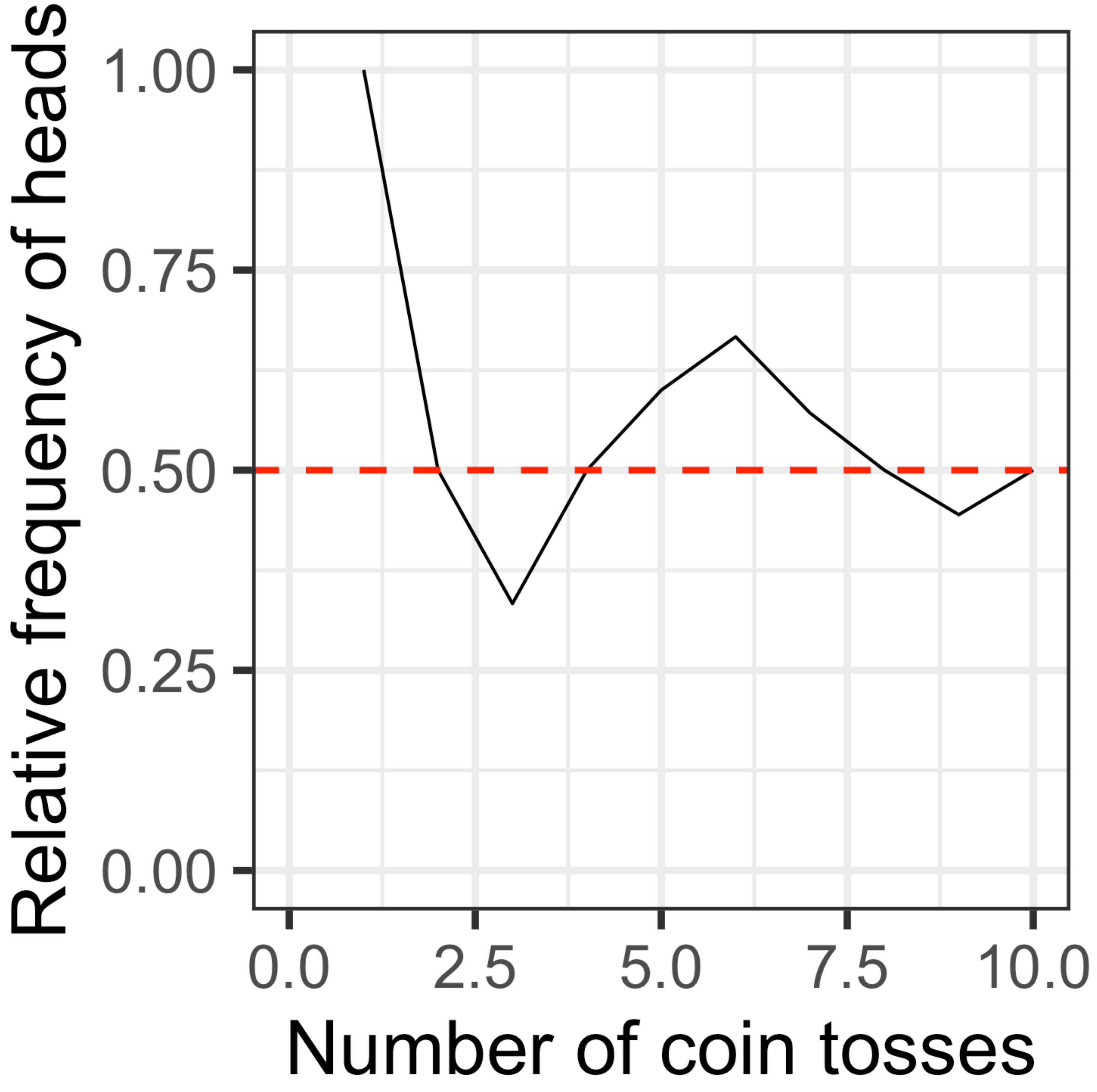
Probability = relative frequency (large sample)

Pr{HEADS} = ?

$$= \frac{\textbf{\# HEADS}}{\textbf{\# Total tosses}}$$

Pr{HEADS} = 0.5

	1/1	1
	1/2	0.5
	1/3	0.33
	2/4	0.5
	3/5	0.6



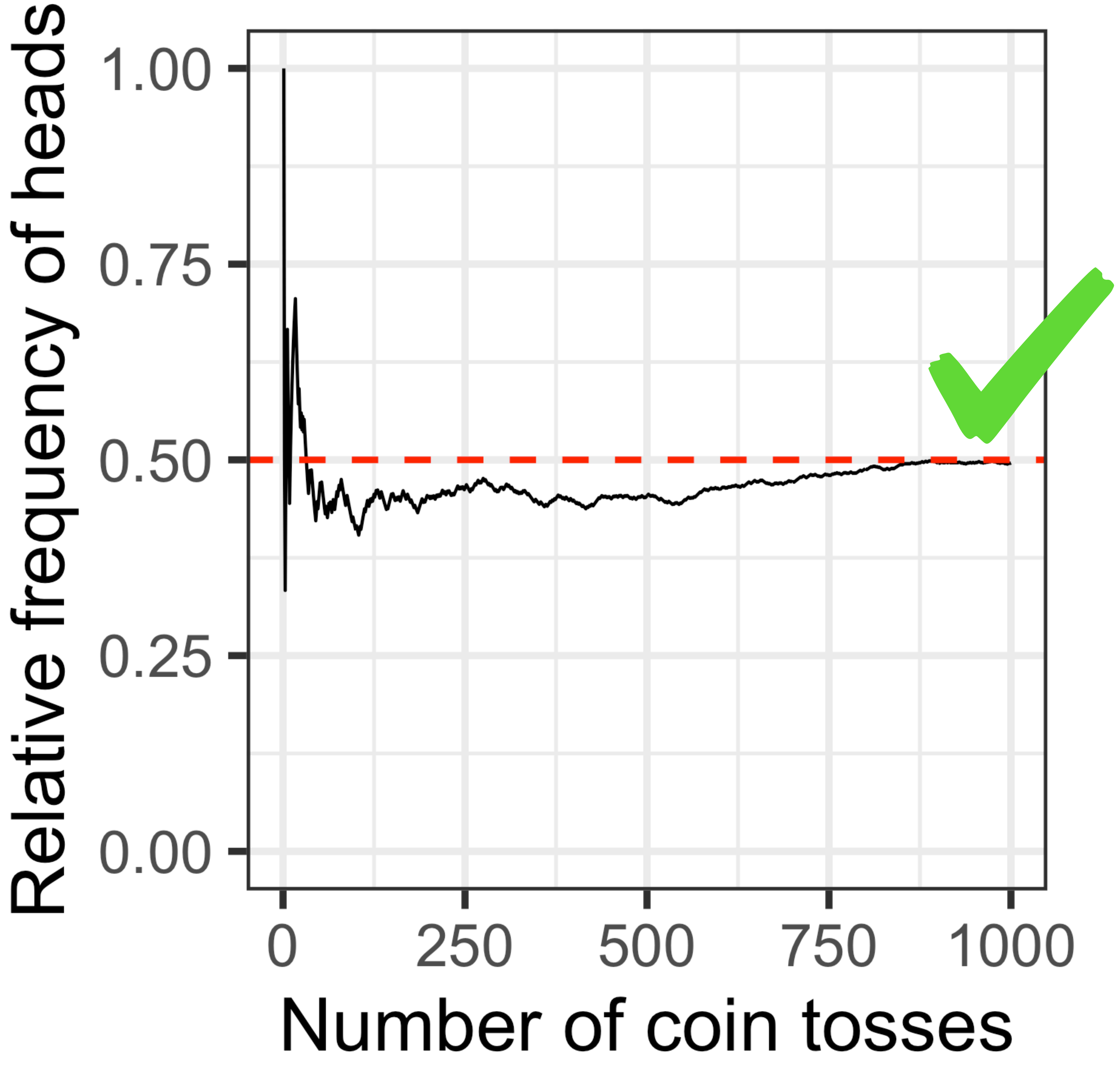
Probability = relative frequency (large sample)

Pr{HEADS} = ?

$$= \frac{\textbf{\# HEADS}}{\textbf{\# Total tosses}}$$

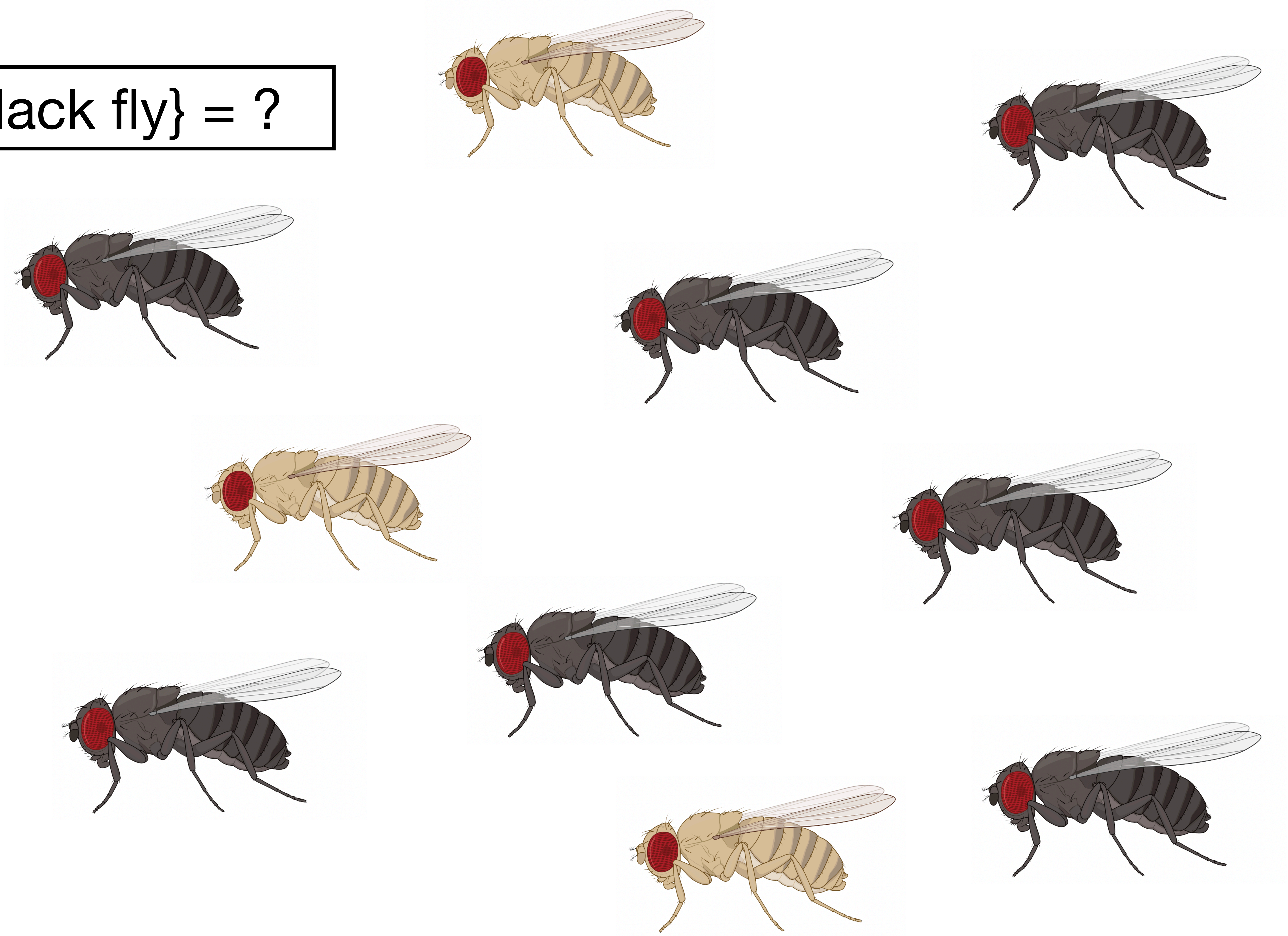
Pr{HEADS} = 0.5

	1/1	1
	1/2	0.5
	1/3	0.33
…	…	…
	501/ 1000	0.5



Probability and random sampling

Pr{Selecting a black fly} = ?



7 black flies

3 white flies

Probability and random sampling

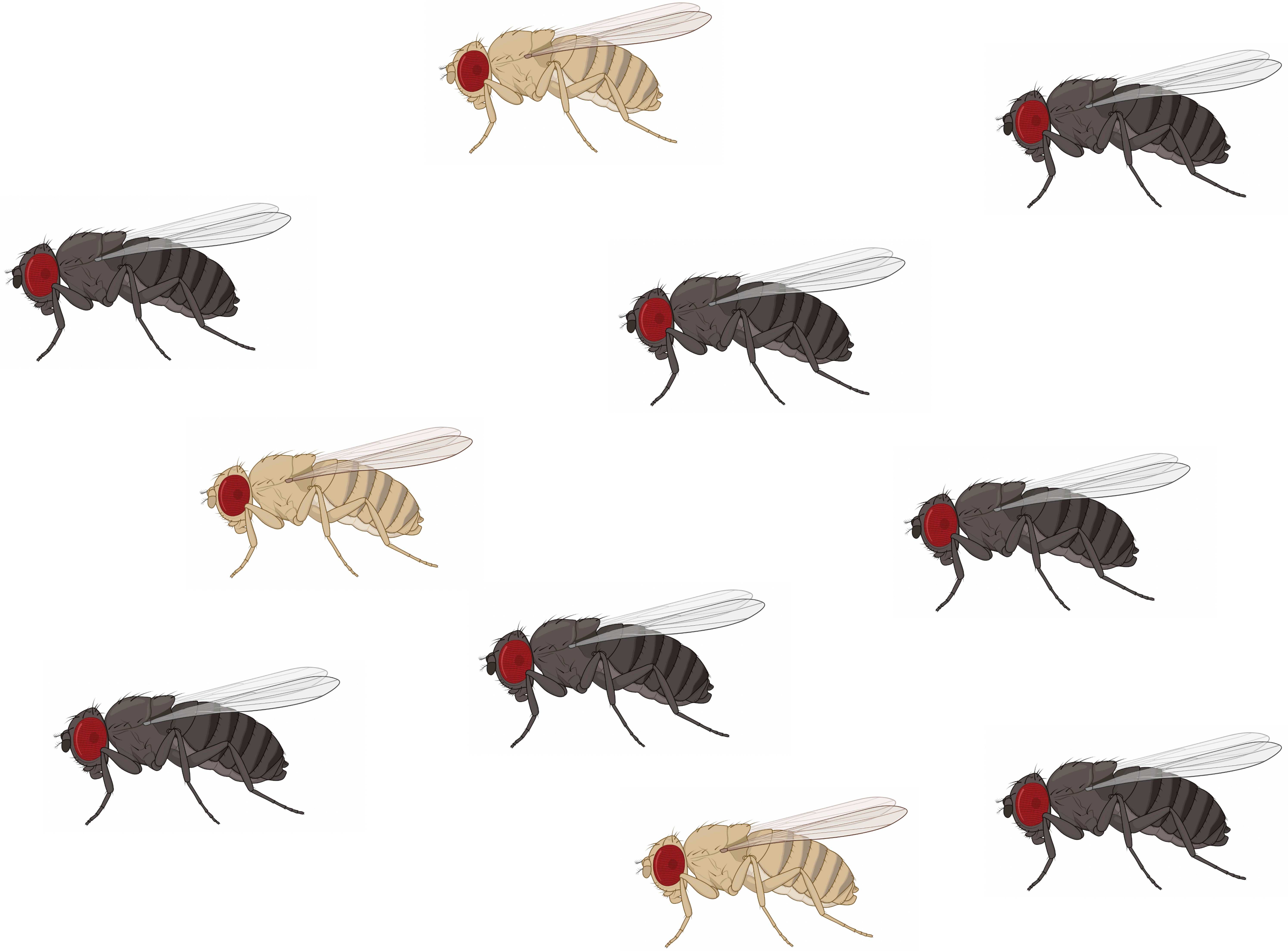
$\boxed{\text{Pr\{Selecting a black fly\} = ?}}$

**assuming random selection*

$$= \frac{\text{\# black flies}}{\text{\# Total flies}}$$

$$= \frac{7}{10}$$

3 white flies; 7 black flies



Rules of probability

- Probability of an event is greater than zero and less than one
- Sum of probabilities of all possible events is equal to one
- Probability that an event, E, does NOT happen is equal to 1 - Pr{E}

Rules of probability

- **Probability of an event is greater than zero and less than one**

Pr{E} = 0	**0 ≤ Pr{E} ≤ 1**	**Pr{E} = 1**
(Never happens)	(Happens with some frequency)	(Always happens)

- Sum of probabilities of all possible events is equal to one

- Probability that an event, E, does NOT happen is equal to 1 - Pr{E}

Rules of probability

- Probability of an event is greater than zero and less than one

$Pr\{E\} = 0$	$0 \leq Pr\{E\} \leq 1$	$Pr\{E\} = 1$
(Never happens)	(Happens with some frequency)	(Always happens)

- **Sum of probabilities of all possible events is equal to one**

$Pr\{H\} = 0.5$ $\qquad$ $Pr\{T\} = 0.5$ $\qquad$ $Pr\{H\} + Pr\{T\} = 1$

- Probability that an event, E, does NOT happen is equal to 1 - Pr{E}

Rules of probability

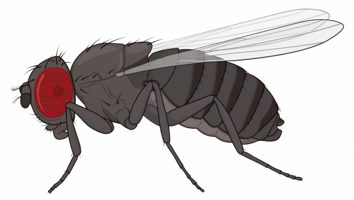
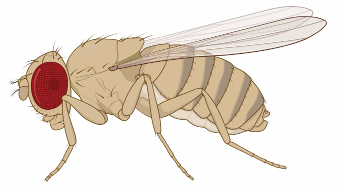
- Probability of an event is greater than zero and less than one

$Pr\{E\} = 0$ (Never happens)	$0 \le Pr\{E\} \le 1$ (Happens with some frequency)	$Pr\{E\} = 1$ (Always happens)

- Sum of probabilities of all possible events is equal to one

  Pr{H} = 0.5 **Pr{T} = 0.5** **Pr{H} + Pr{T} = 1**

- **Probability that an event, E, does NOT happen is equal to 1 - Pr{E}**

7  **;** **3**  **Pr{W} = 1 - Pr{B}** = 1 - 7/10 = 3/10

Rules of probability

- **<u>Probability of an event is greater than zero and less than one</u>**

Pr{E} = 0
(Never happens)

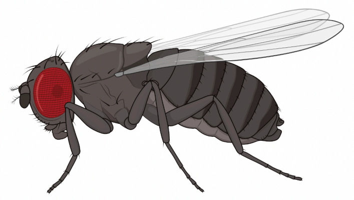
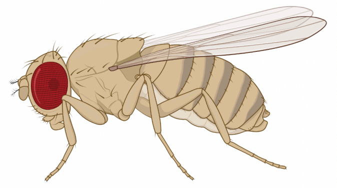
0 ≤ Pr{E} ≤ 1
(Happens with some frequency)

Pr{E} = 1
(Always happens)

- **<u>Sum of probabilities of all possible events is equal to one</u>**


Pr{H} = 0.5 **Pr{T} = 0.5** **Pr{H} + Pr{T} = 1**

- **<u>Probability that an event, E, does NOT happen is equal to 1 - Pr{E}</u>**

7  **;** **3** 

Pr{W} = 1 - Pr{B} = 1 - 7/10 = 3/10

Table 1 shows the distribution of ages of Americans. Find the probability that the age of a randomly chosen American is between 20 and 49.

Age	Relative freq
0-19	0.27
20-29	0.14
30-39	0.13
40-49	0.14
50-64	0.19
65+	0.13

- *Probability ~= relative frequency*
- *Sum(probs) = 1*

Table 1 shows the distribution of ages of Americans. Find the probability that the age of a randomly chosen American is between 20 and 49.

Age	Probability
0-19	0.27
20-29	0.14
30-39	0.13
40-49	0.14
50-64	0.19
65+	0.13

- *Probability ~= relative frequency*
- *Sum(probs) = 1*

$$\boxed{0.27 + 0.14 + 0.13 + 0.14 + 0.19 + 0.13 = 1}$$

$$\Pr\{20\text{-}49\} = \Pr\{20s\} + \Pr\{30s\} + \Pr\{40s\}$$

$$= 0.14 + 0.13 + 0.14$$

$$\mathbf{= 0.41}$$

Table 1 shows the distribution of ages of Americans. Find the probability that the age of a randomly chosen American is less than 65.

Age	Probability
0-19	0.27
20-29	0.14
30-39	0.13
40-49	0.14
50-64	0.19
65+	0.13

- *Probability ~= relative frequency*
- *Sum(probs) = 1*

$$\Pr\{<65\} = \Pr\{<20\} + \Pr\{20s\} + \Pr\{30s\} + \Pr\{40s\} + \Pr\{50s\}$$

$$\Pr\{<65\} = 1 - \Pr\{65+\}$$

$$= 1 - 0.13$$

$$\mathbf{= 0.87}$$

Probability for continuous variables?

Discrete

Continuous

Categorical

Blood type (A, B, AB, O)

Fish sex (male, female)

Shape of pea (smooth, wrinkled)

Success in trial (Alive, dead)

Numeric

Human height

Blood cholesterol of patient

Number of bacterial colonies

Length of DNA segment

Discrete

Probability for continuous variables?

Discrete

Continuous

<u>Categorical</u>

Blood type (A, B, AB, O)

Fish sex (male, female)

Shape of pea (smooth, wrinkled)

Success in trial (Alive, dead)

<u>Numeric</u>

Human height

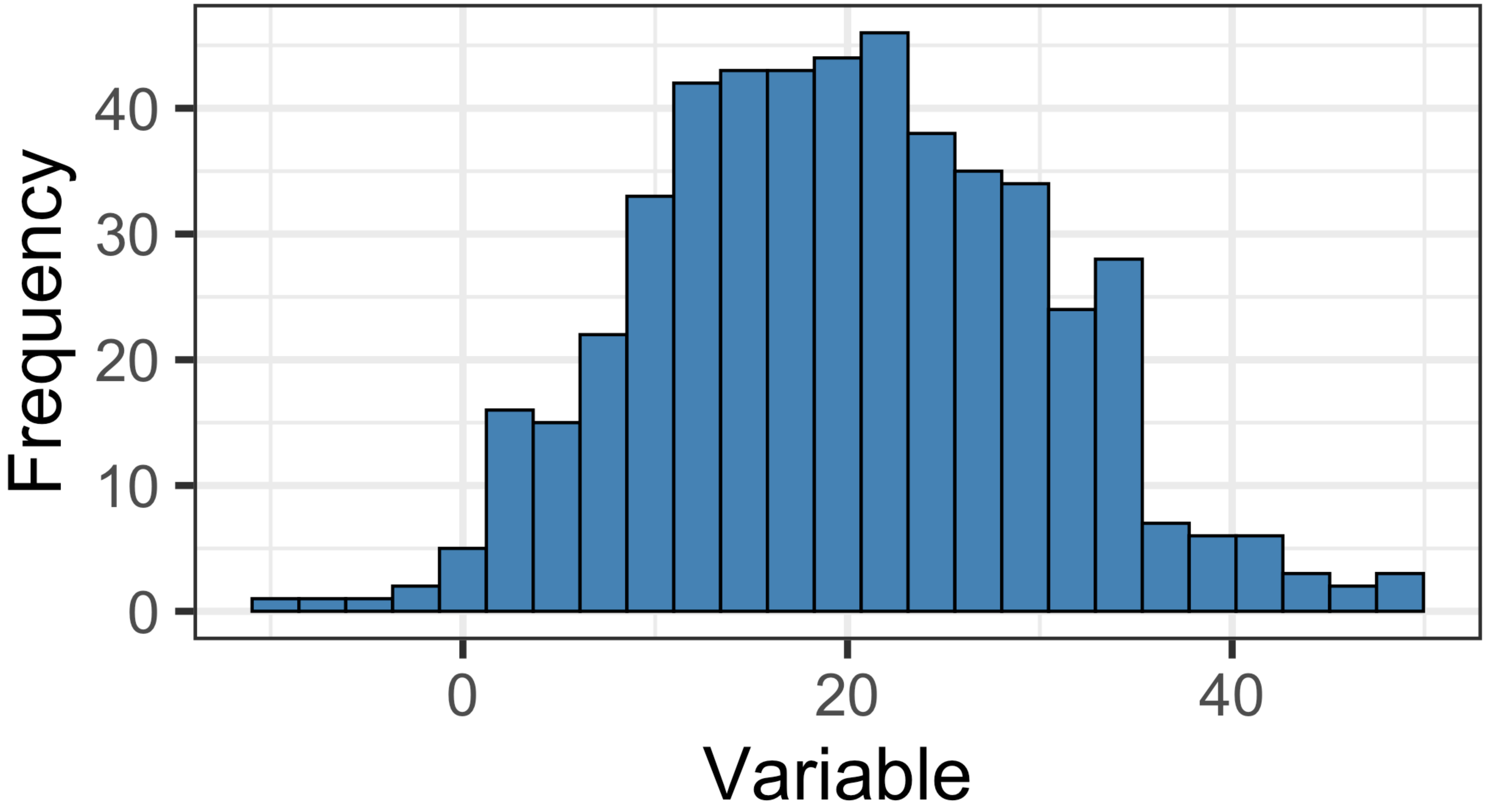
Blood cholesterol of patient

Number of bacterial colonies

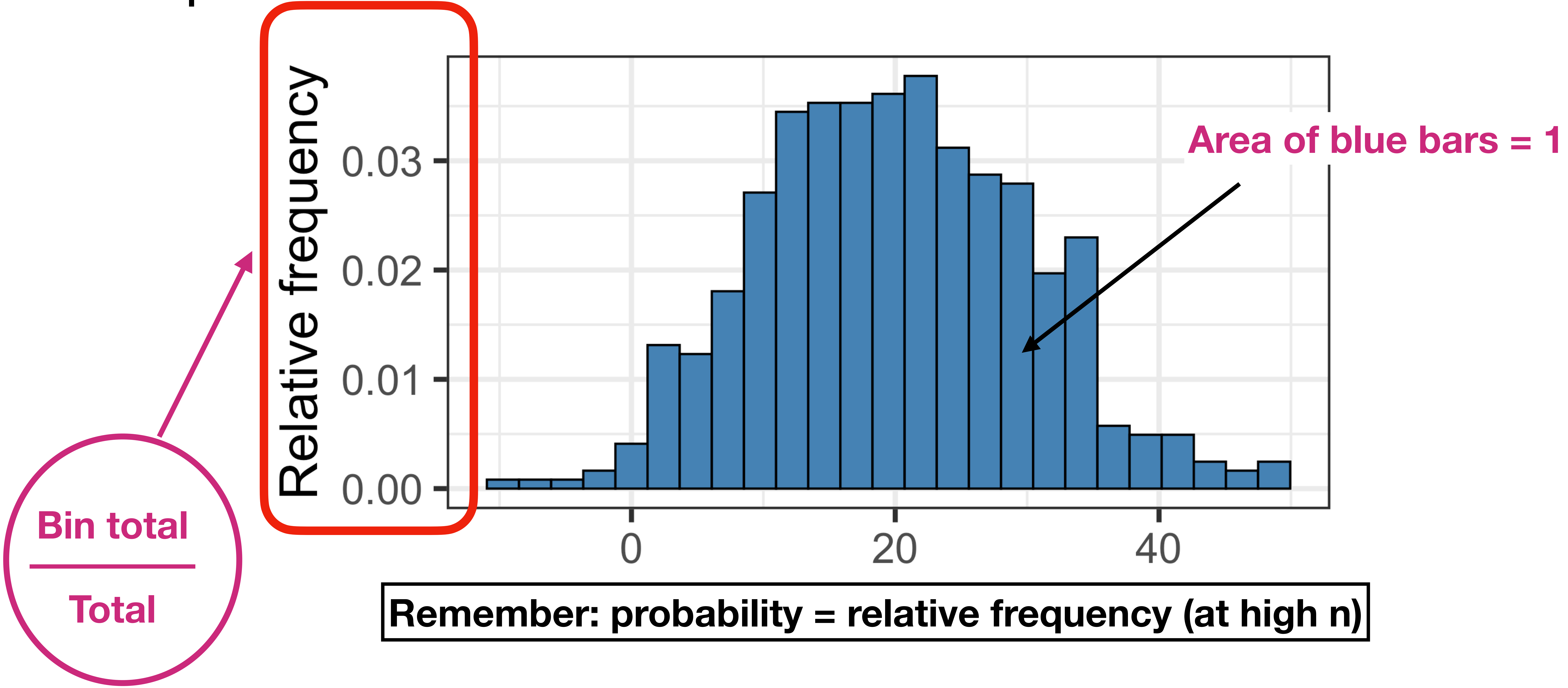
Length of DNA segment

Discrete

Probability for continuous variables?

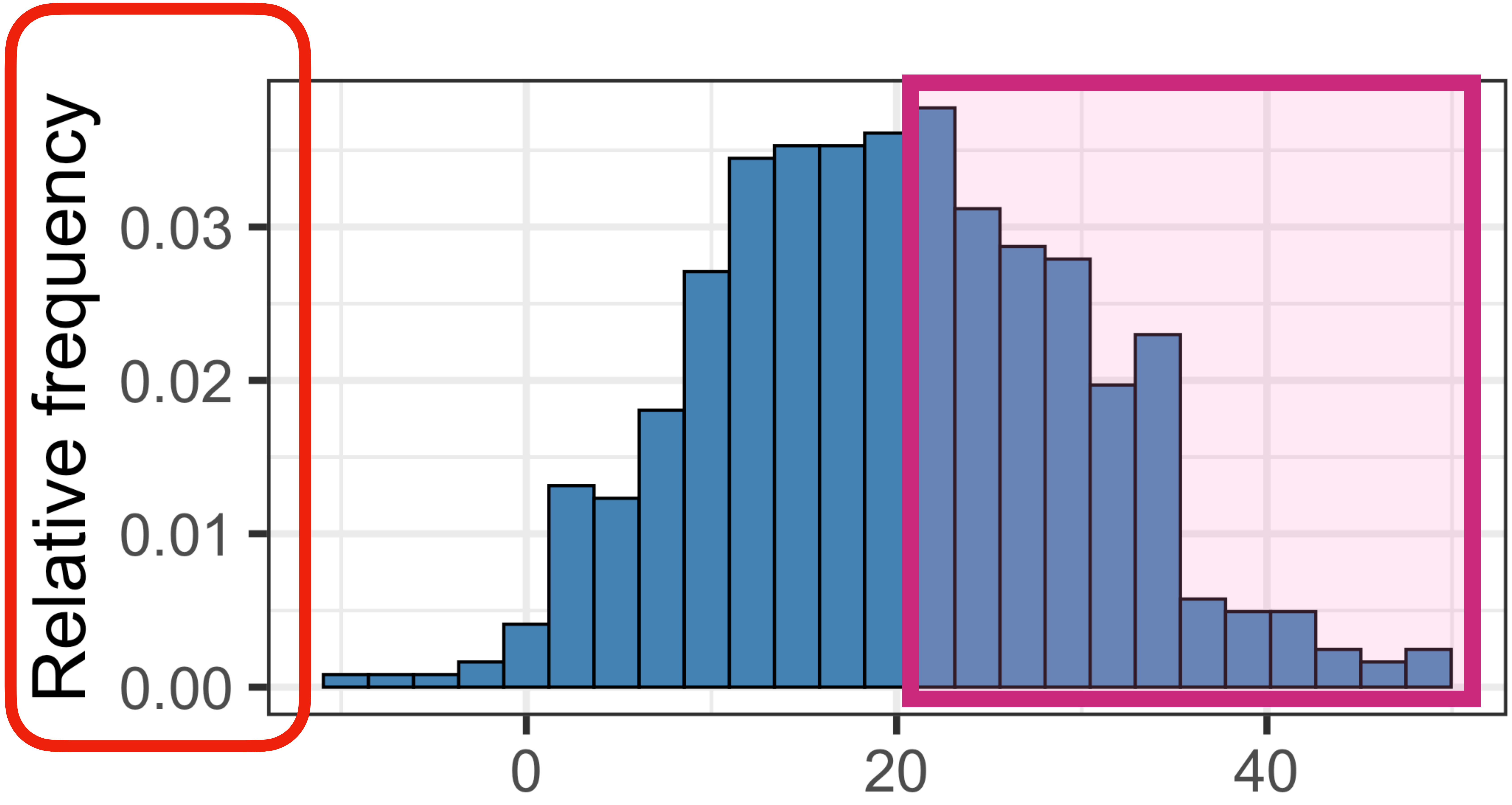


Relative frequency plots are used to estimate probabilities for continuous variables



Remember: probability = relative frequency (at high n)

Relative frequency plots are used to estimate probabilities for continuous variables



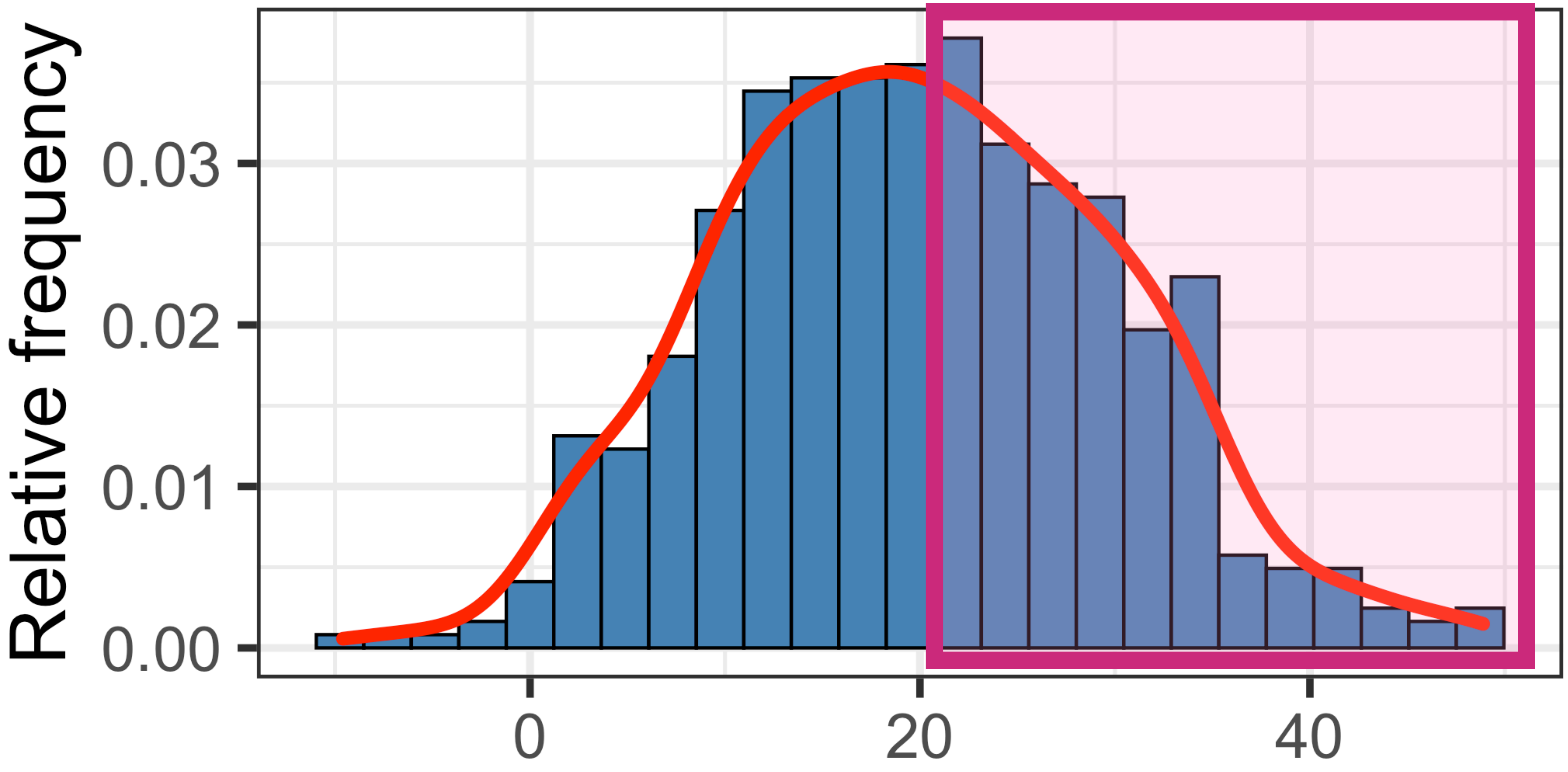
What is the probability the value > 20?

Add up the relative frequencies of all the bins with values > 20!

Estimating ~ 50%

Remember: probability = relative frequency (at high n)

Density curves are used to estimate probabilities for continuous variables



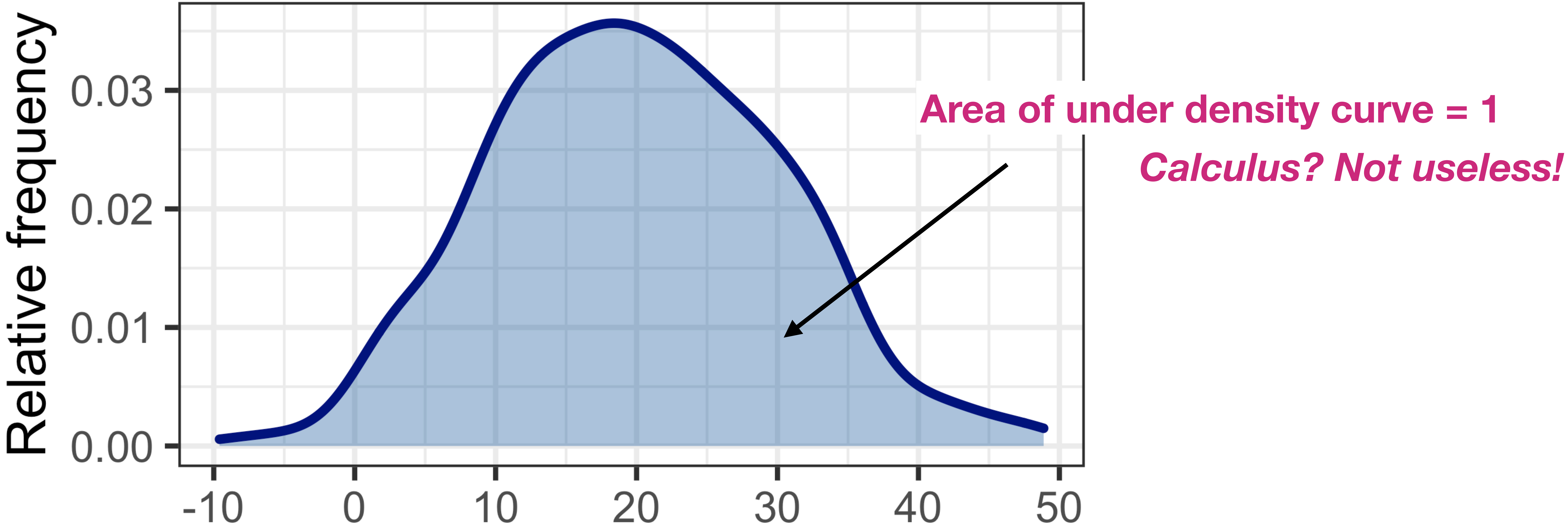
What is the probability the value > 20?

Add up the relative frequencies of all the bins with values > 20!

Estimating ~ 50%

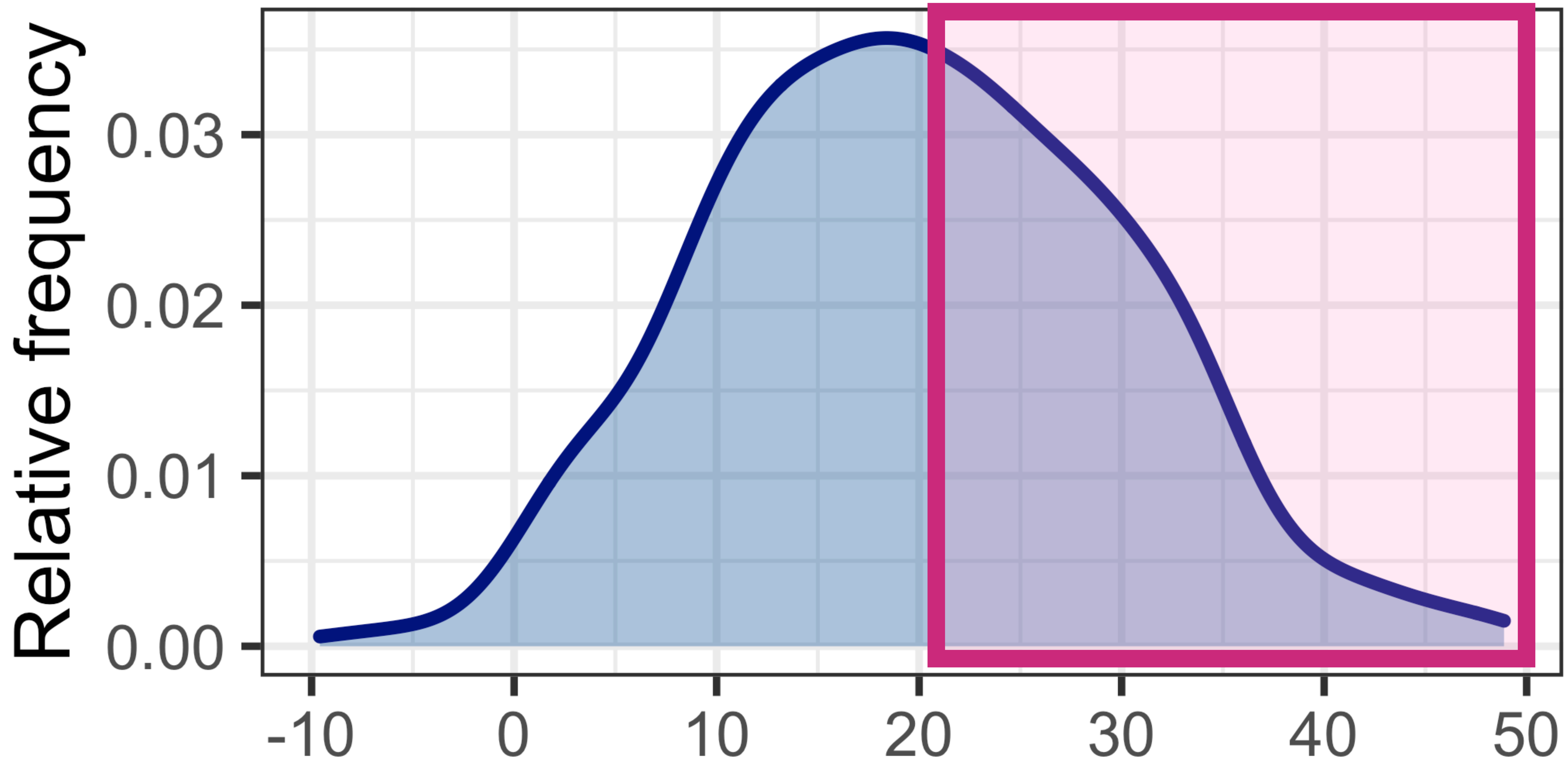
Remember: probability = relative frequency (at high n)

Density curves are used to estimate probabilities for continuous variables



Remember: probability = relative frequency (at high n)

Density curves are used to estimate probabilities for continuous variables



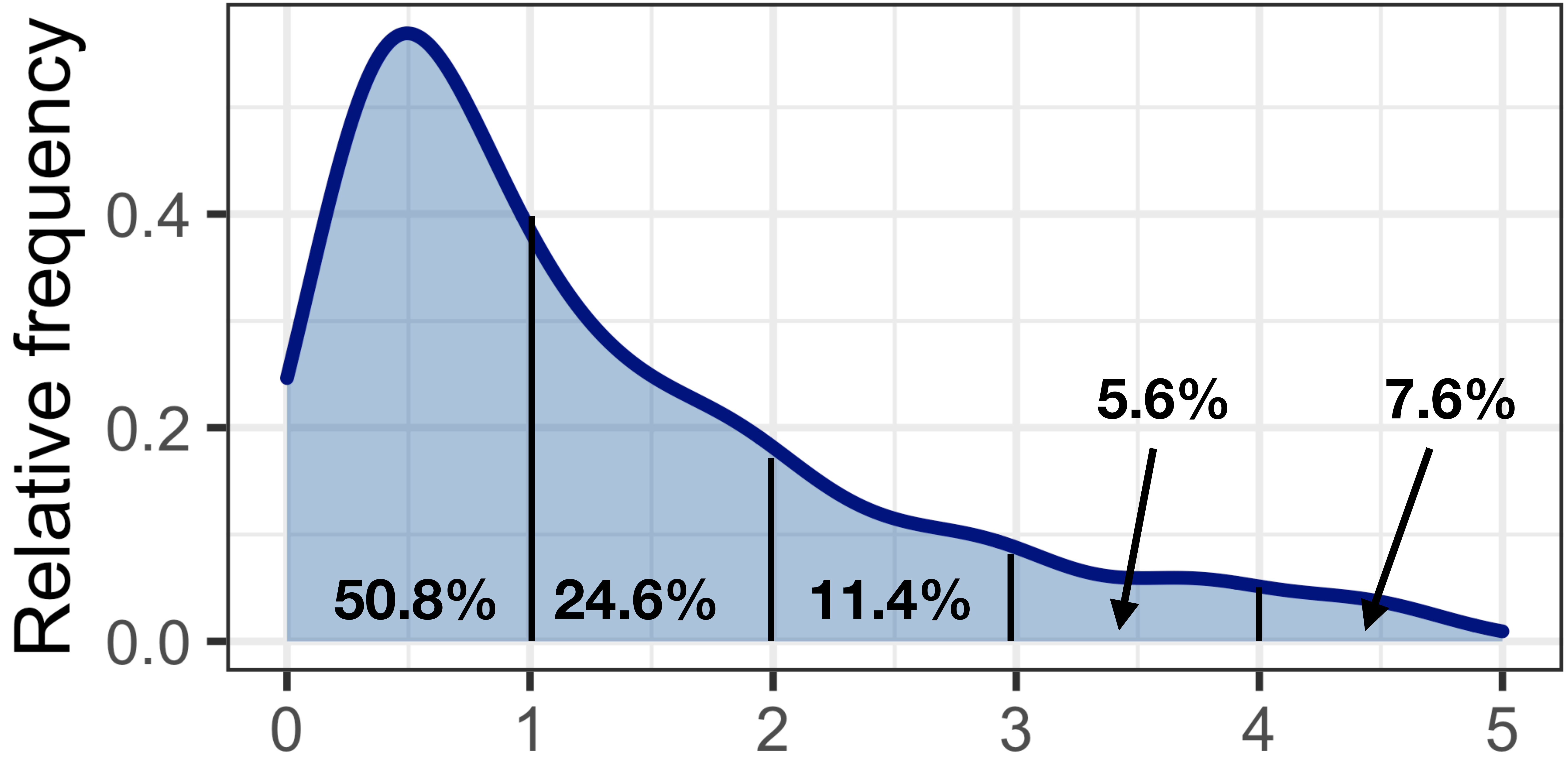
Remember: probability = relative frequency (at high n)

What is the probability the value > 20?

Find the area under the curve between 20 and max

Estimating ~ 50%

Density curves are used to estimate probabilities for continuous variables



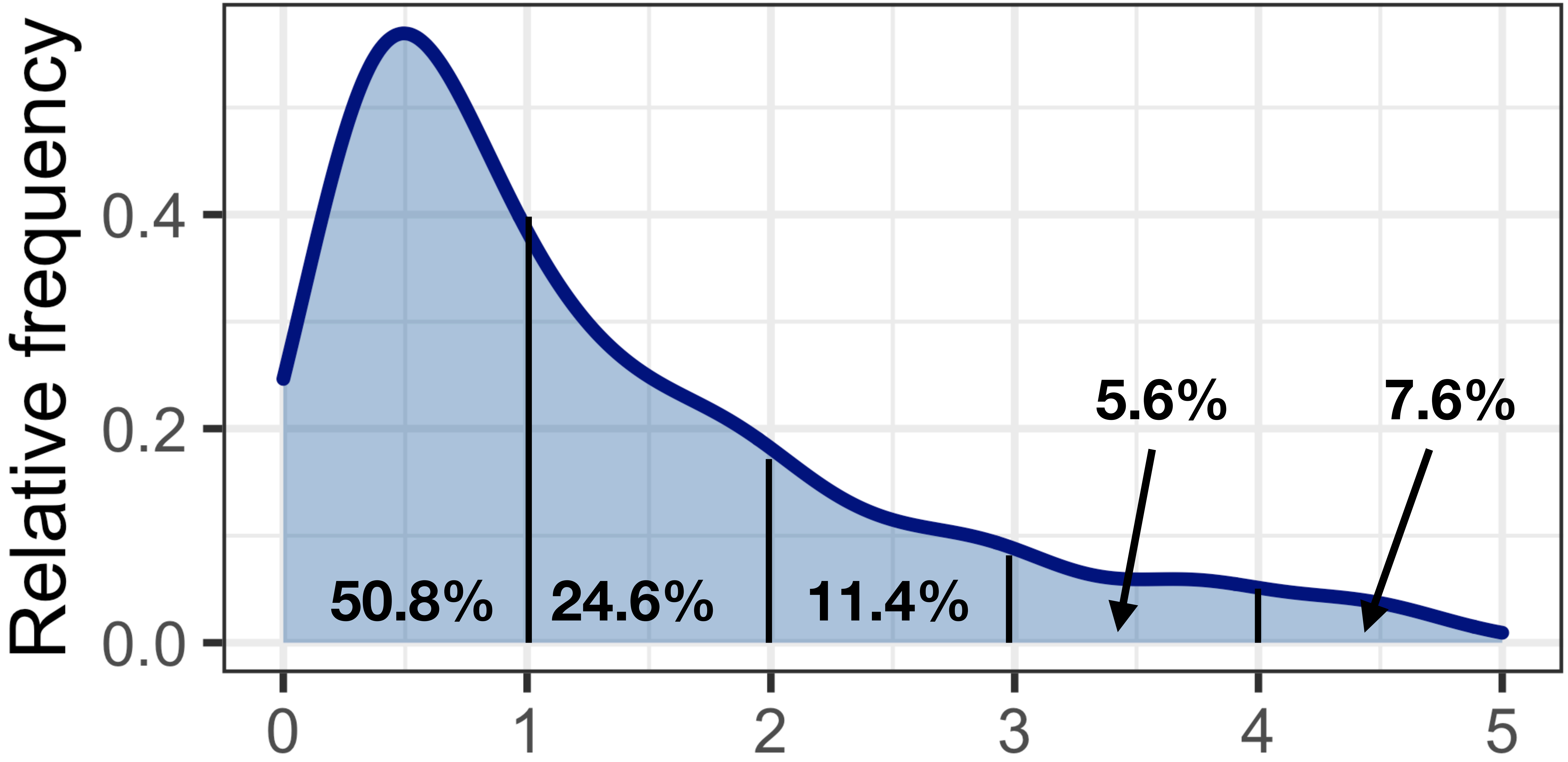
What is the probability that a given value falls between 1 and 3?

Add up the proportion of values between 1 and 3

24.6 + 11.4 = **37.8%**

Remember: probability = relative frequency (at high n)

Density curves are used to estimate probabilities for continuous variables



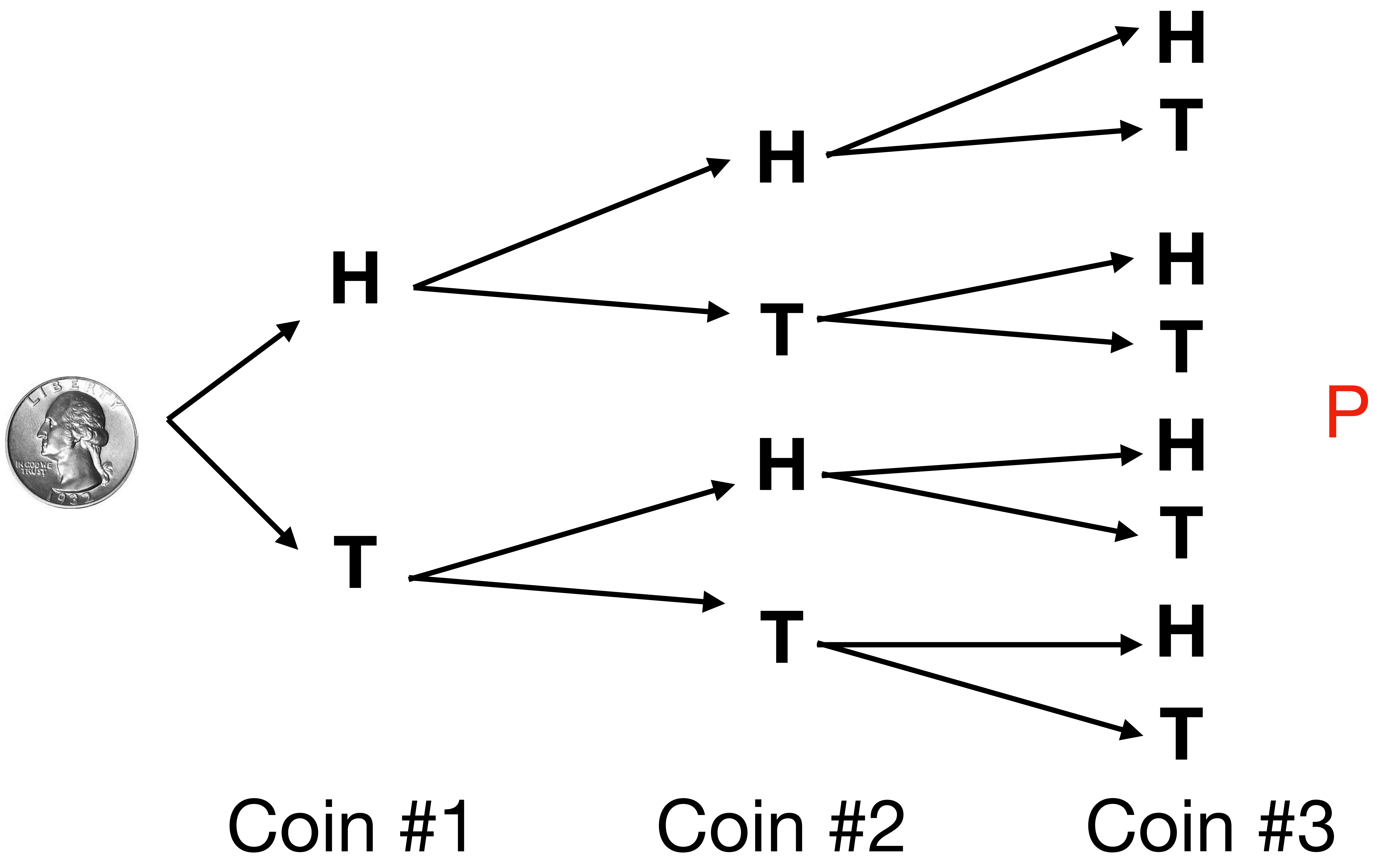
Remember: probability = relative frequency (at high n)

Suppose you take a sample of two individuals, what is the probability that both values will be > 1?

TBD…

Visualizing probabilities with trees

What is the probability of flipping three heads in a row?

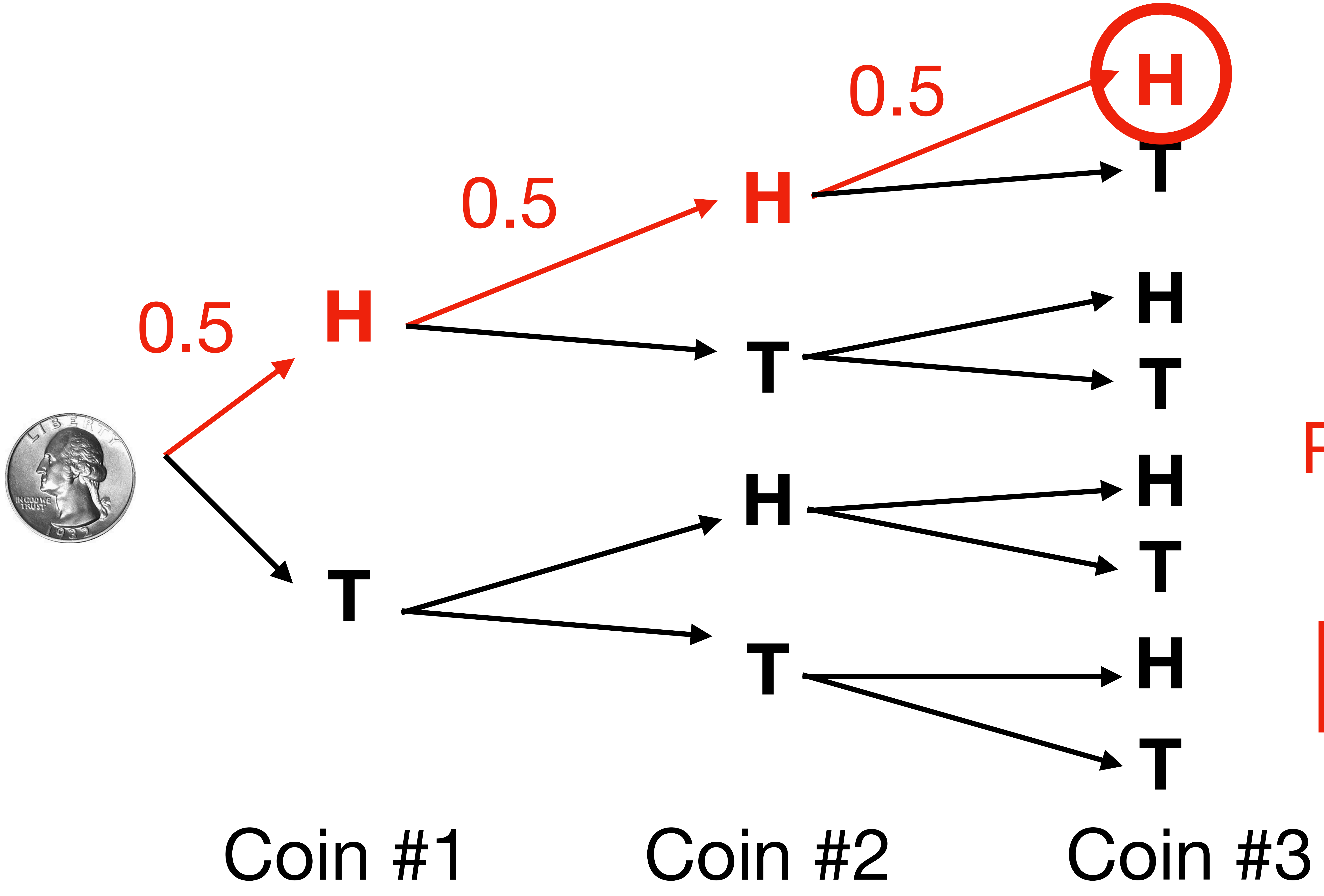


- *Probability ~= relative frequency*
- *Sum(probs) = 1*

$$\Pr = \frac{\text{\# Fav. outcomes}}{\text{\# Total outcomes}}$$

Visualizing probabilities with trees

What is the probability of flipping three heads in a row?



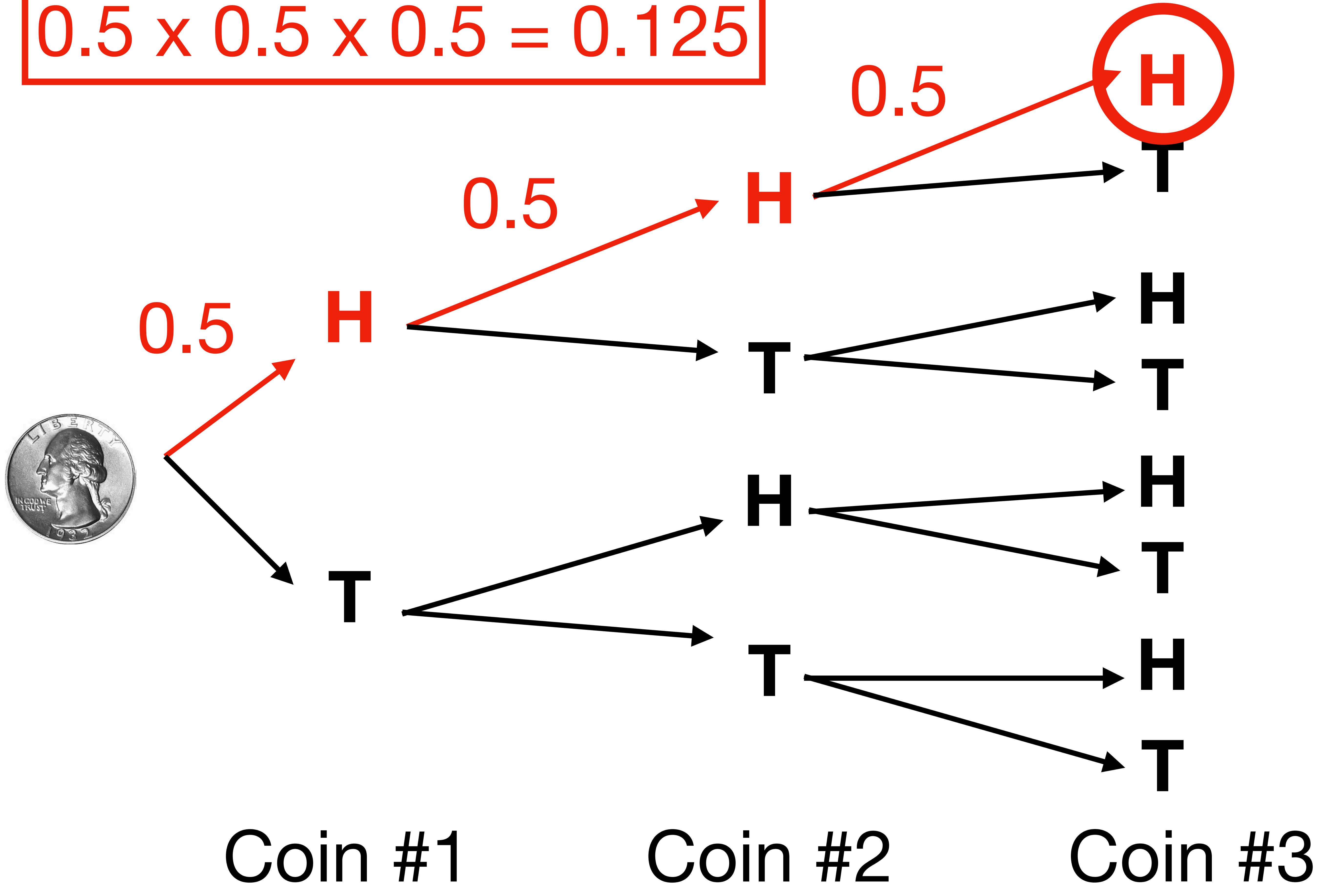
- *Probability ~= relative frequency*
- *Sum(probs) = 1*

$$\Pr = \frac{\text{\# Fav. outcomes}}{\text{\# Total outcomes}} = \frac{\mathbf{1}}{\mathbf{8}}$$

$$\boxed{0.5 \times 0.5 \times 0.5 = 0.125 \text{ (or } 1/8\text{)}}$$

Visualizing probabilities with trees

$\boxed{0.5 \times 0.5 \times 0.5 = 0.125}$

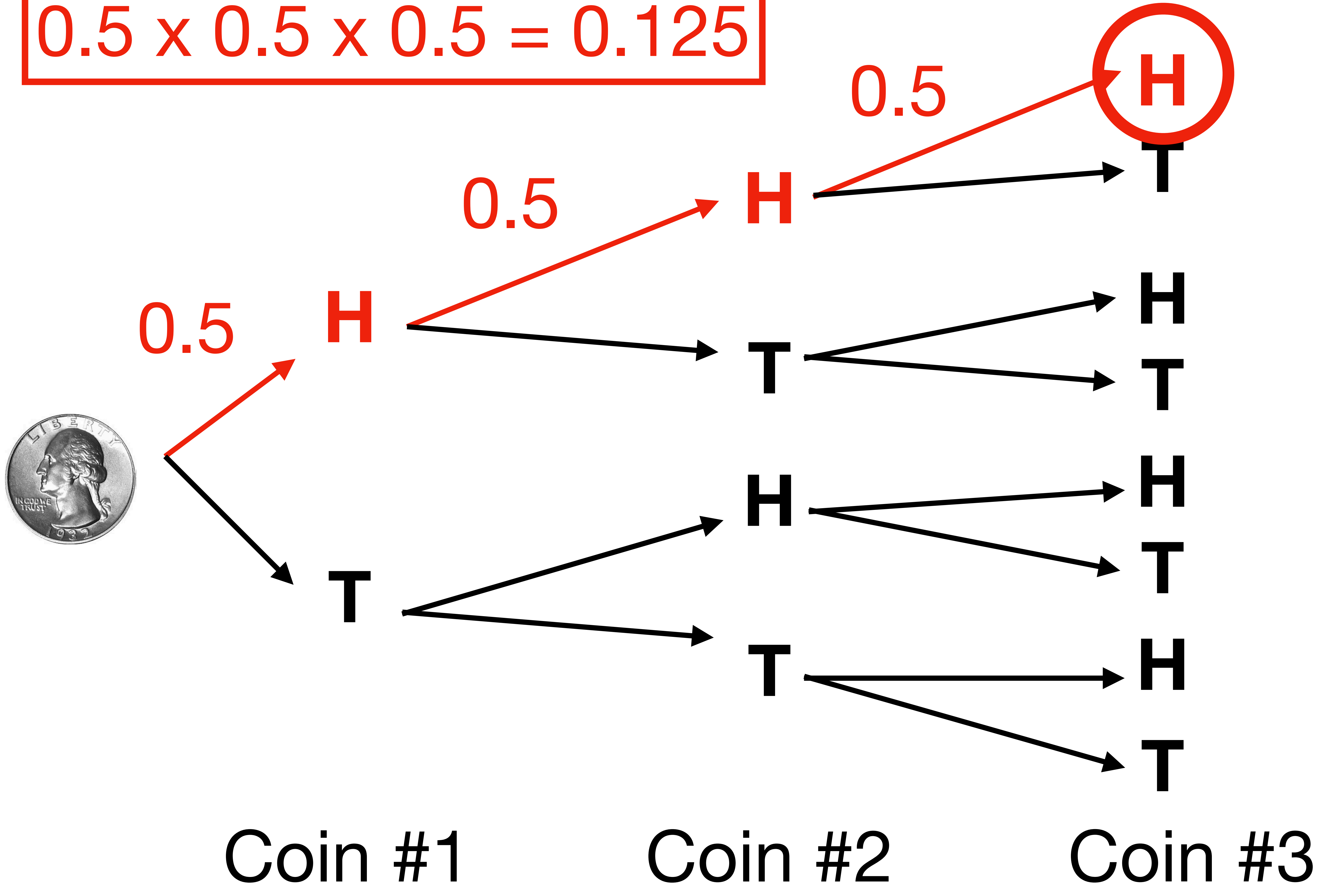


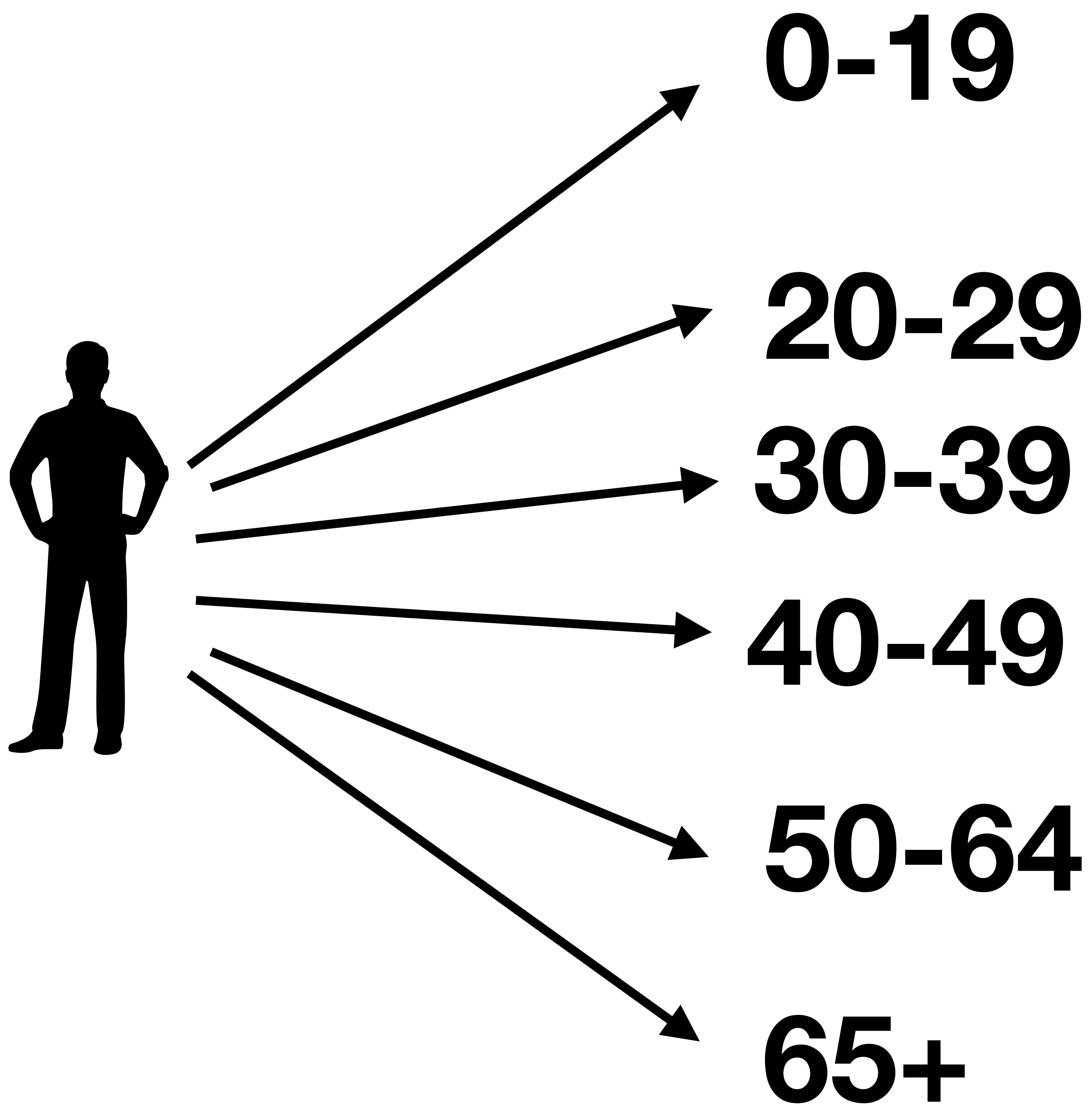
Age	Probability
0-19	0.27
20-29	0.14
30-39	0.13
40-49	0.14
50-64	0.19
65+	0.13

$\boxed{0.14 + 0.13 + 0.14 = 0.41}$

Visualizing probabilities with trees

0.5 x 0.5 x 0.5 = 0.125

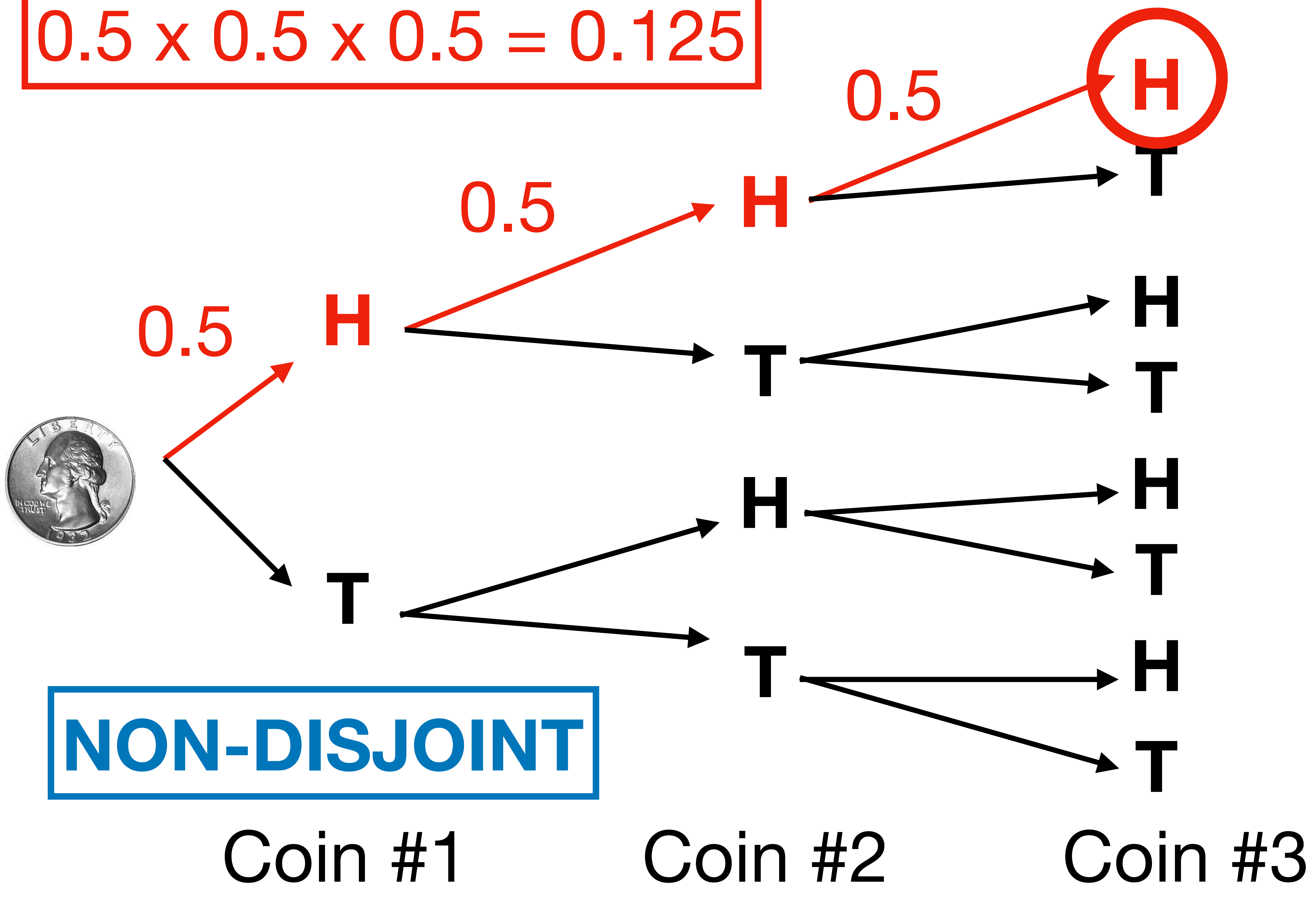




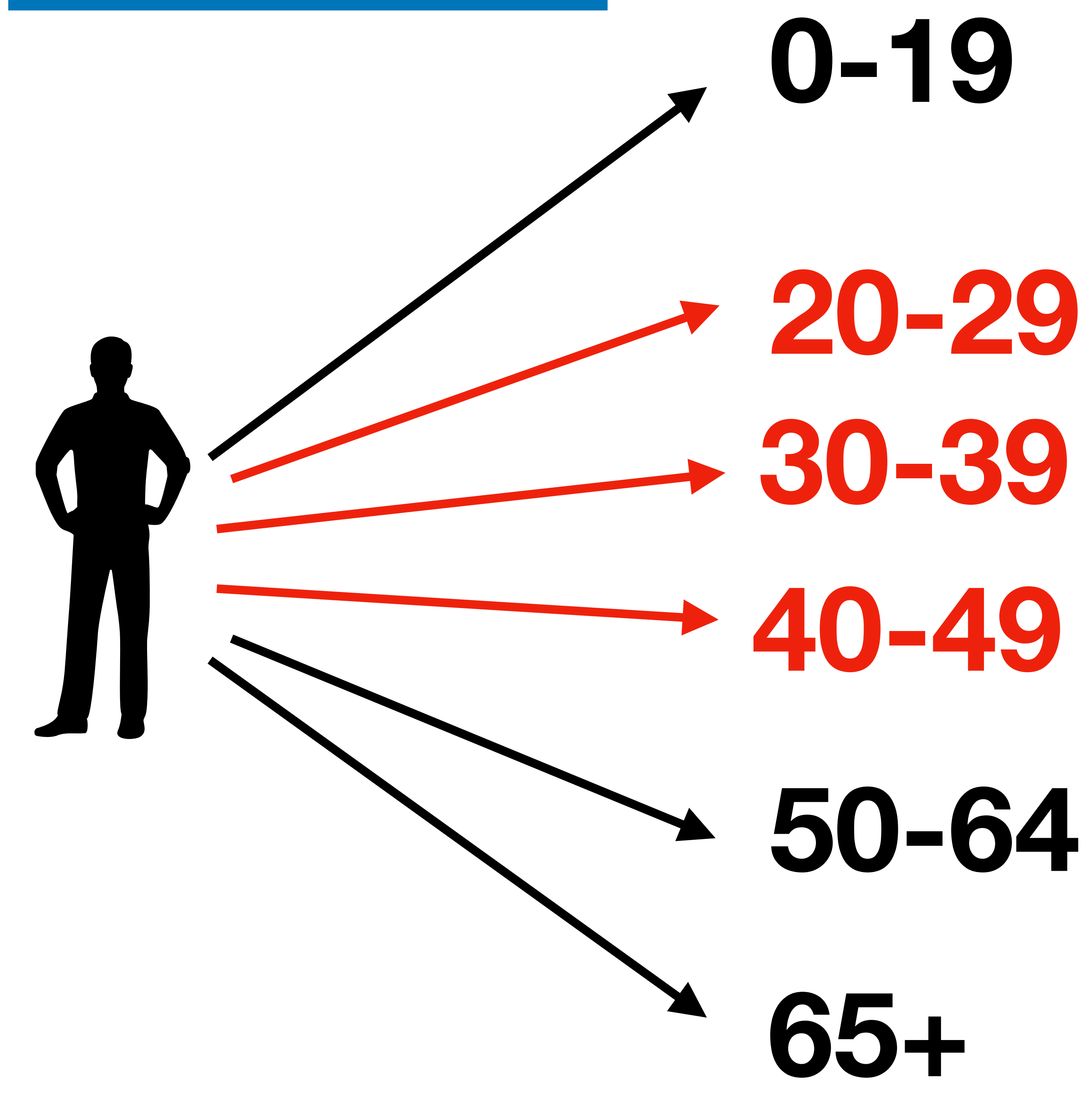
0.14 + 0.13 + 0.14 = 0.41

Visualizing probabilities with trees

0.5 x 0.5 x 0.5 = 0.125

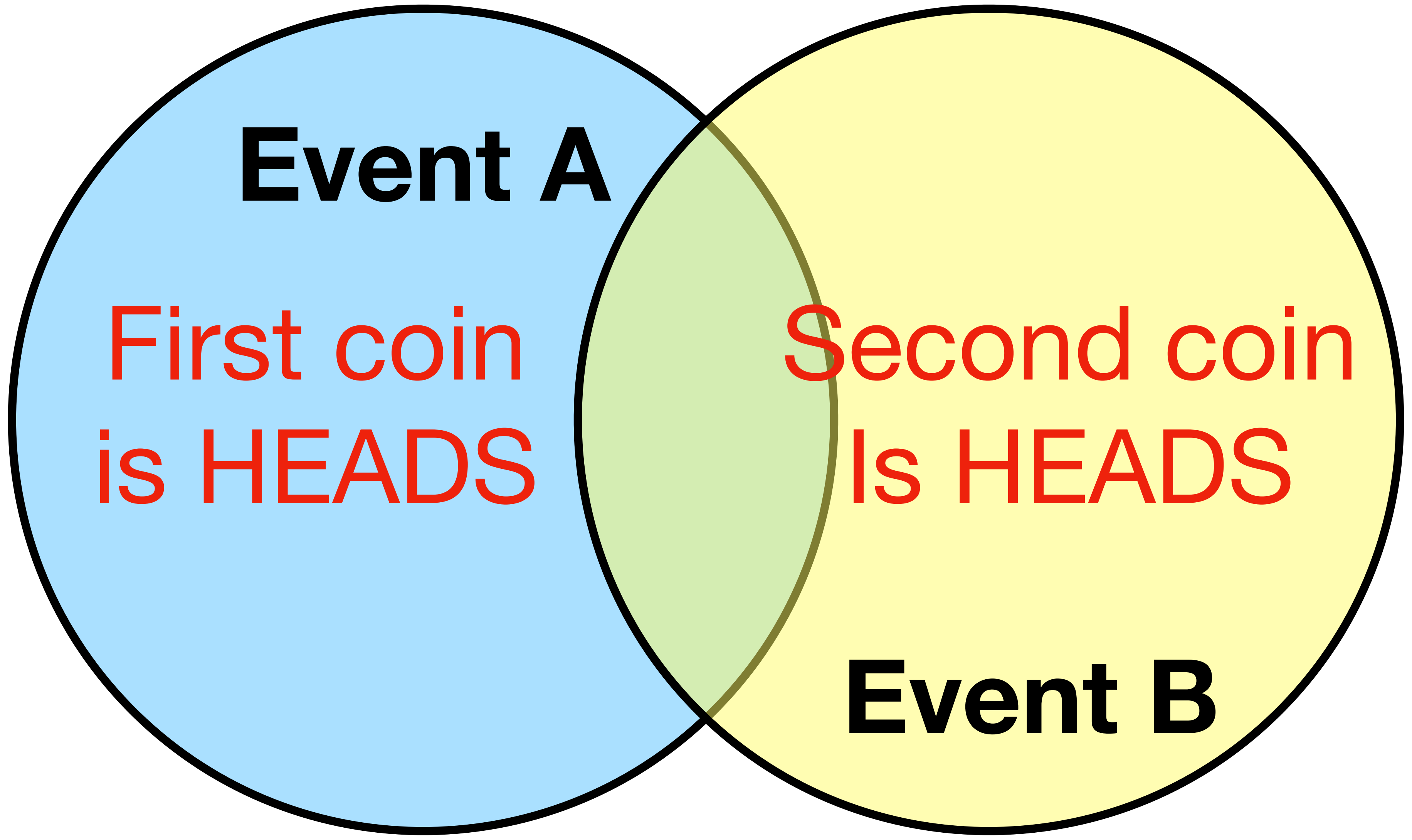


DISJOINT *(Mutually exclusive)*

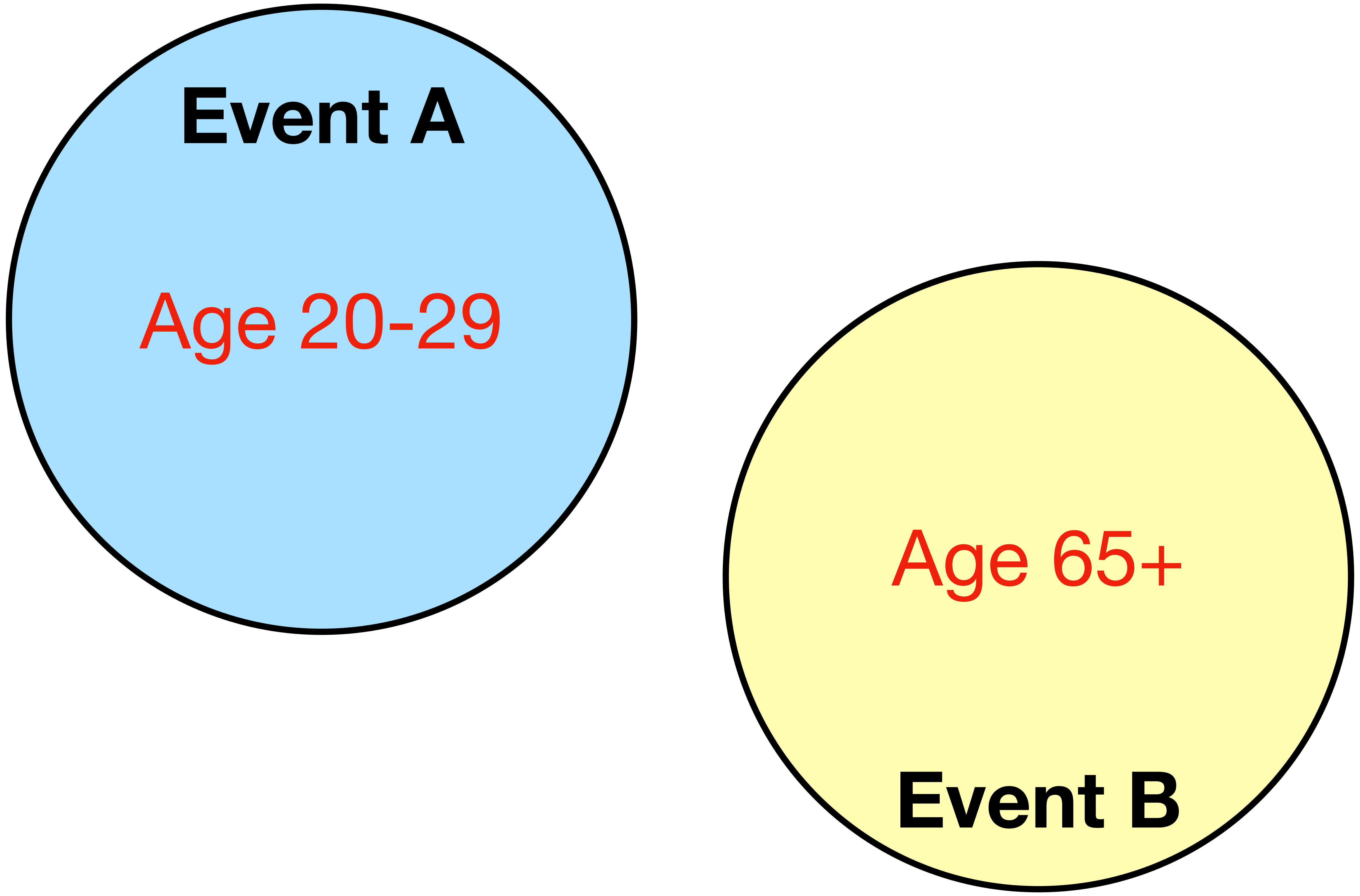


0.14 + 0.13 + 0.14 = 0.41

Defining **disjoint** events (mutually exclusive)

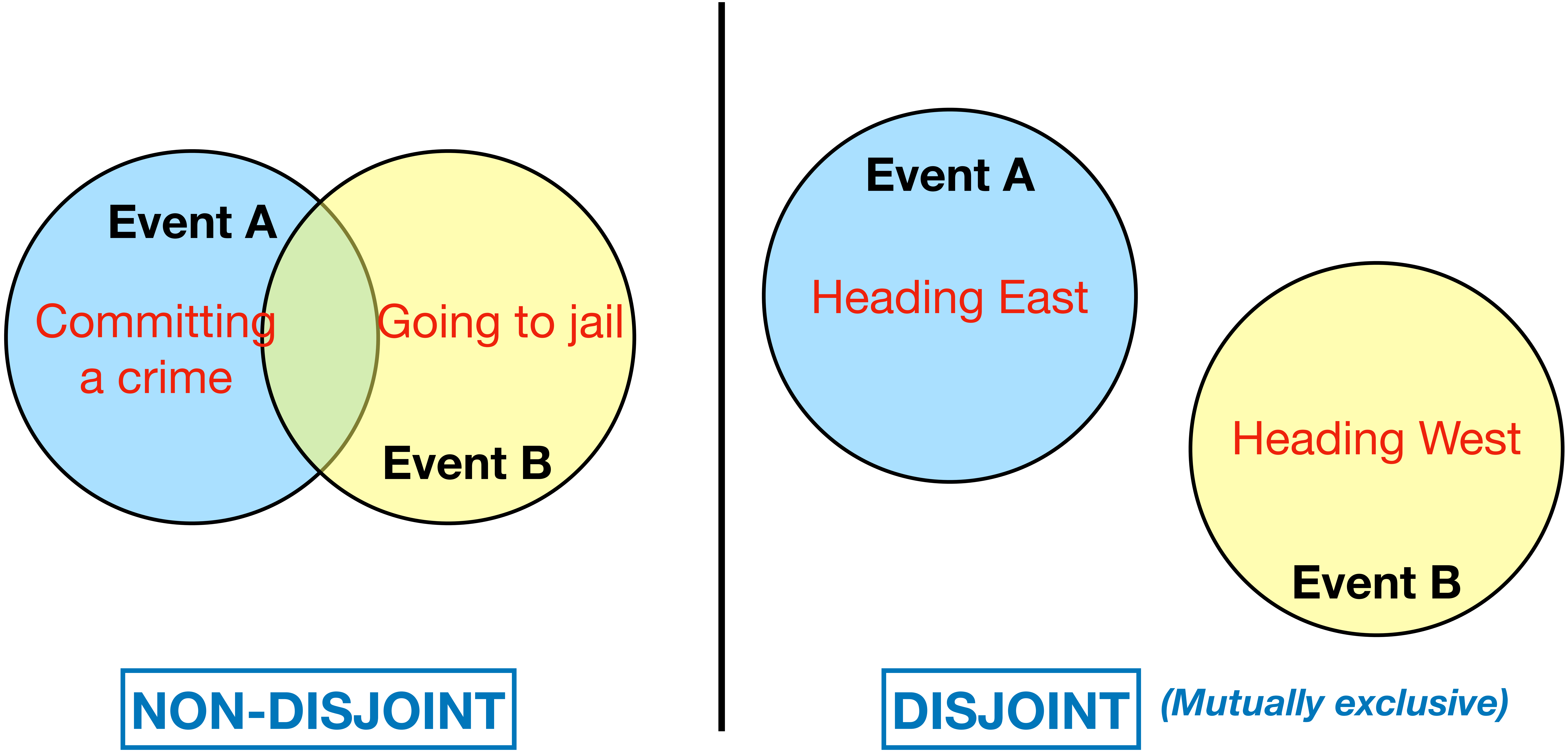


NON-DISJOINT

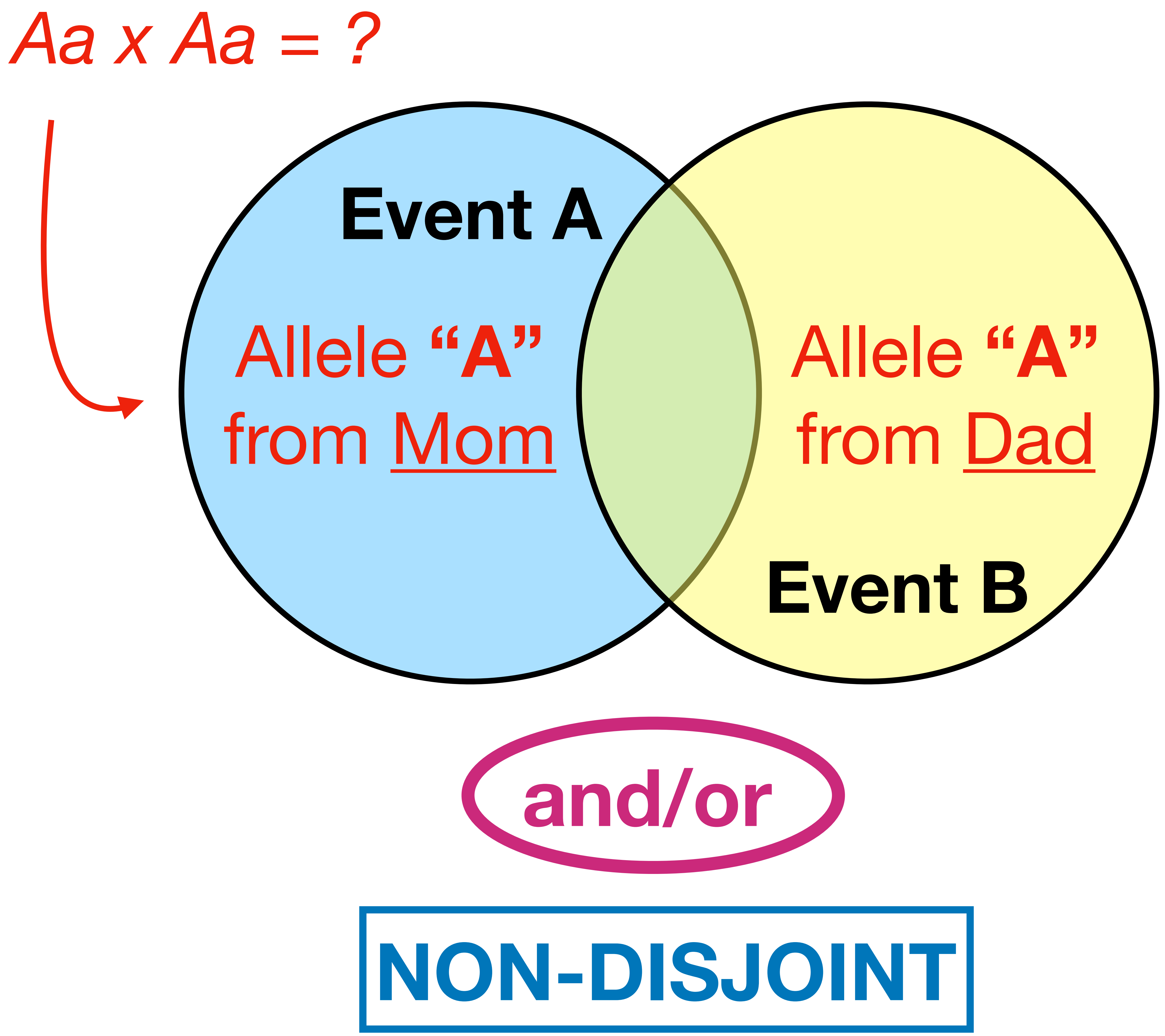


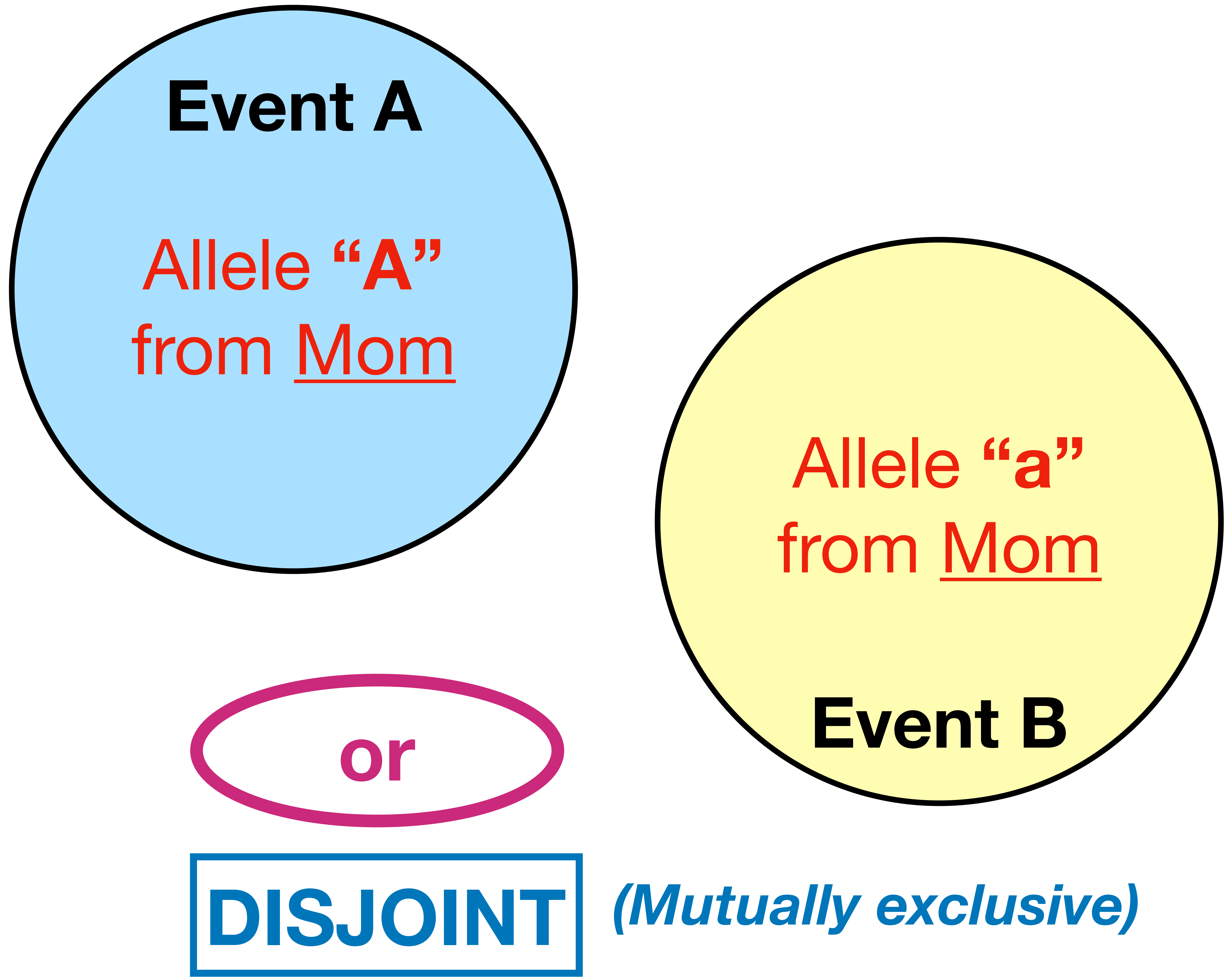
DISJOINT ***(Mutually exclusive)***

Defining **disjoint** events (mutually exclusive)



Defining **disjoint** events (mutually exclusive)

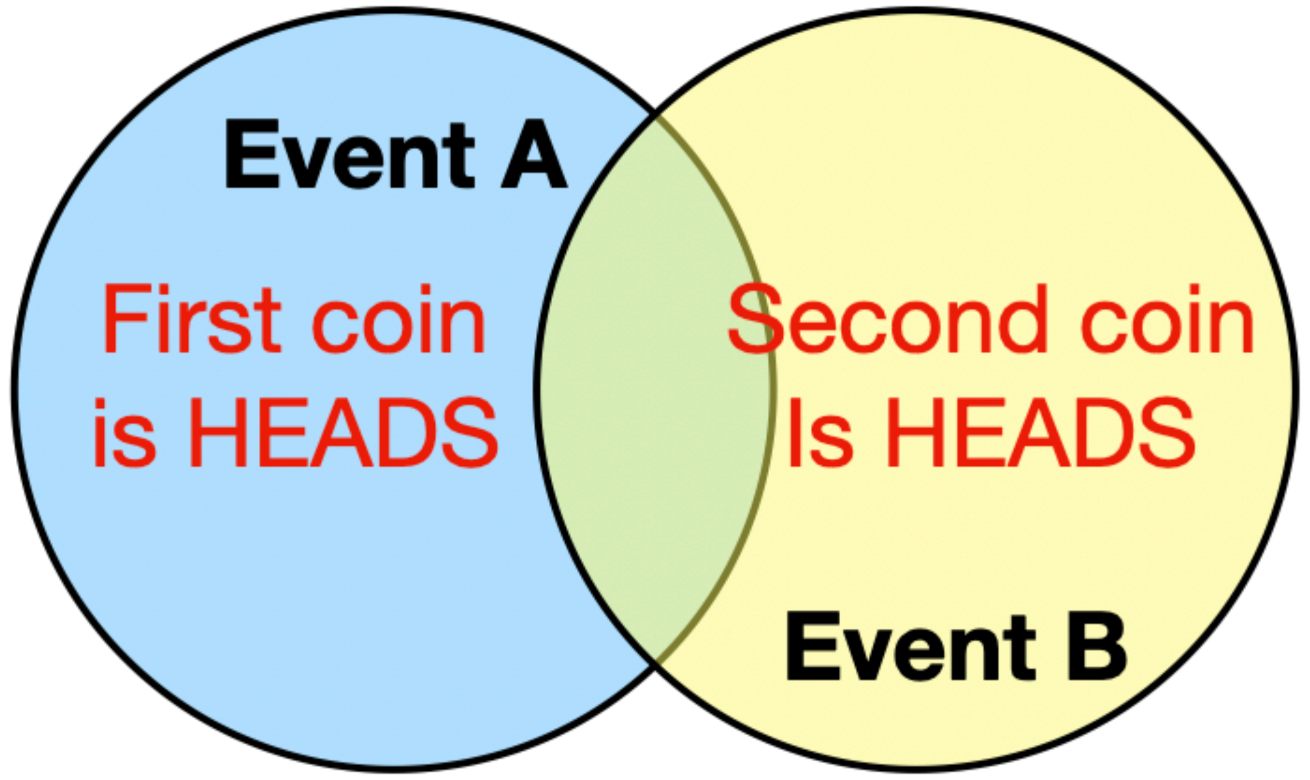




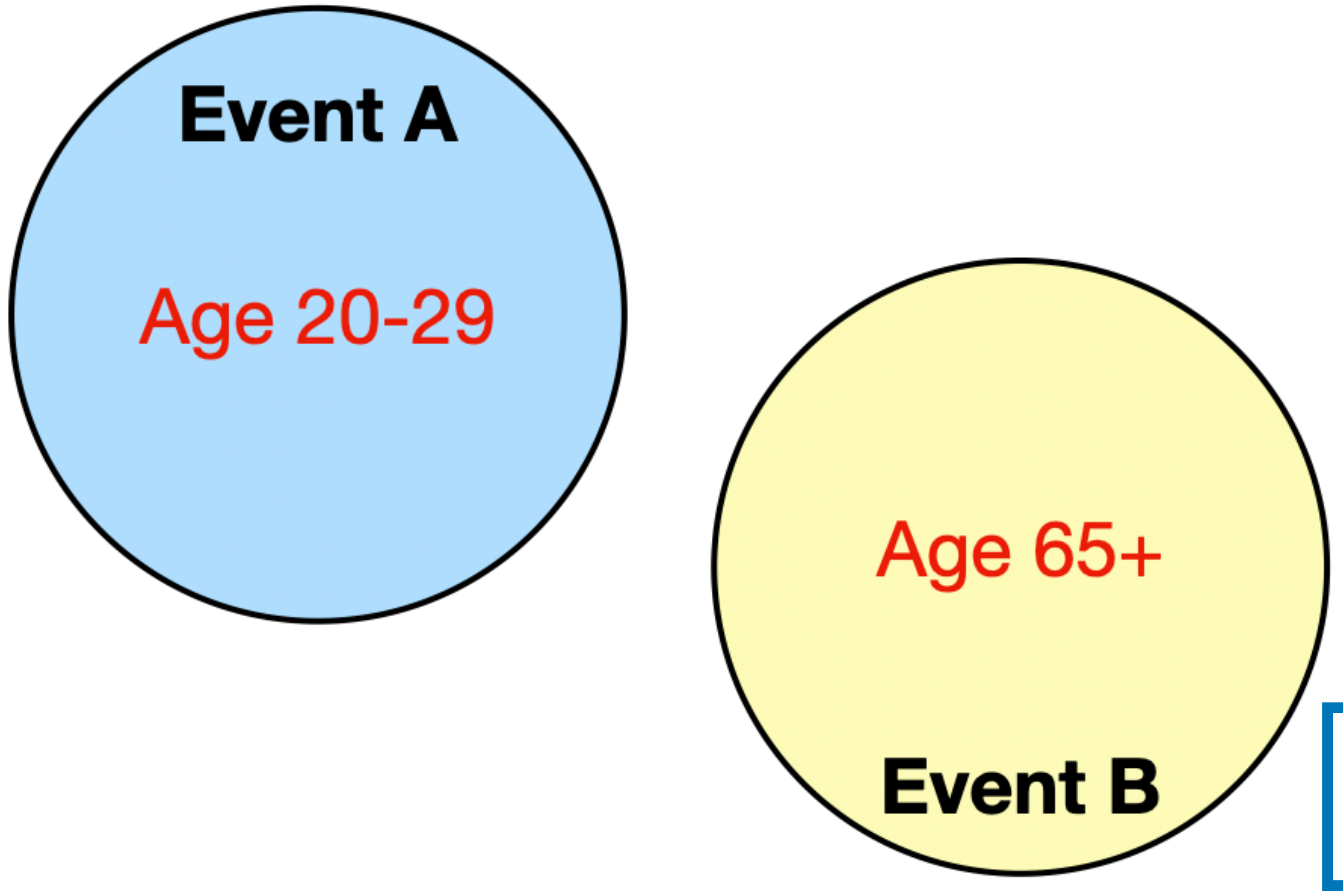
Defining **independent** events

- Two events are **independent** if the fact that one event has occurred **does not affect** the probability that the other event will occur

- Two events are **dependent** if the fact that one event has occurred **does affect** the probability that the other event will occur



Independent



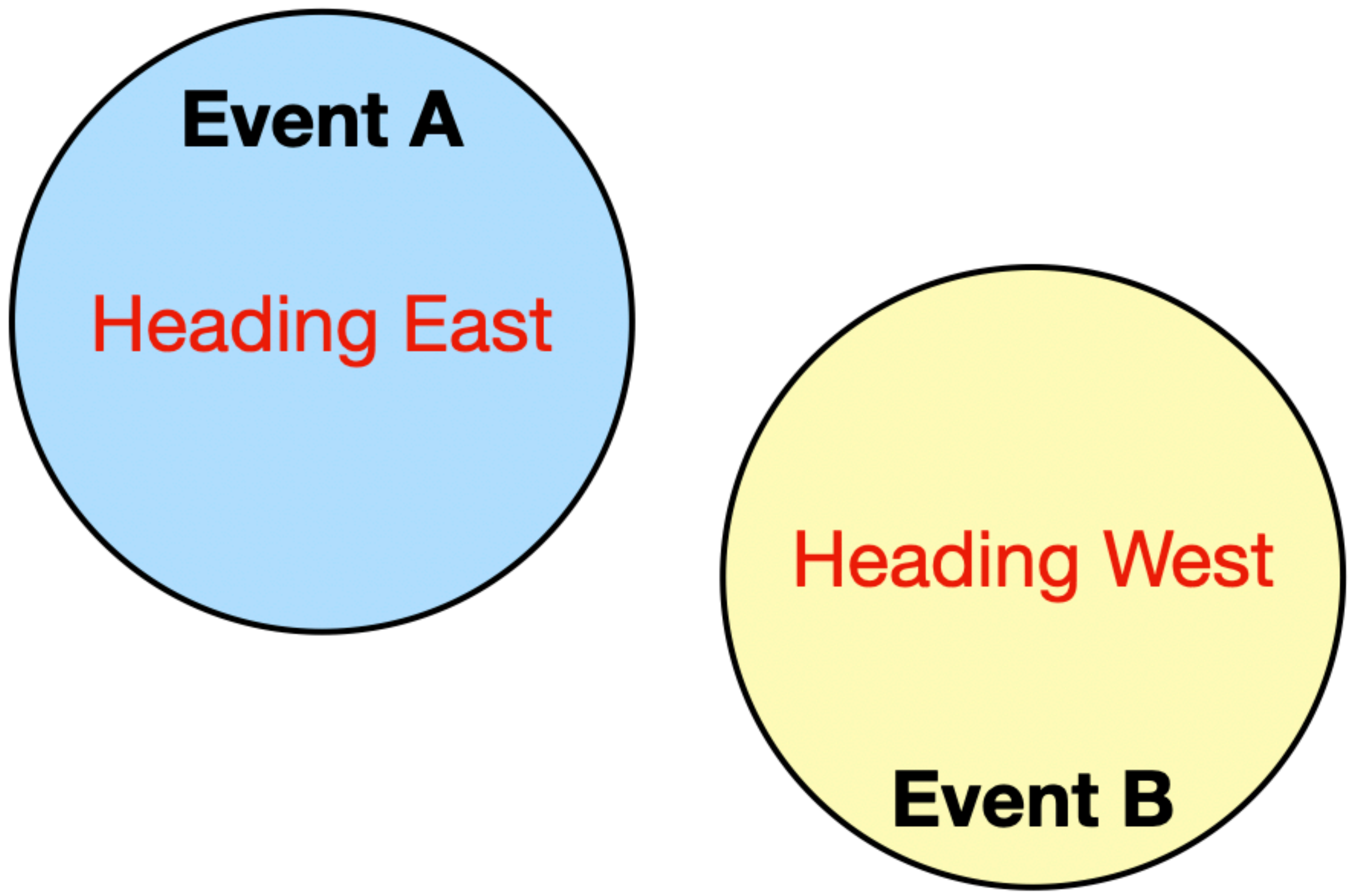
Dependent

Defining **independent** events

- Two events are **independent** if the fact that one event has occurred **does not affect** the probability that the other event will occur

- Two events are **dependent** if the fact that one event has occurred **does affect** the probability that the other event will occur

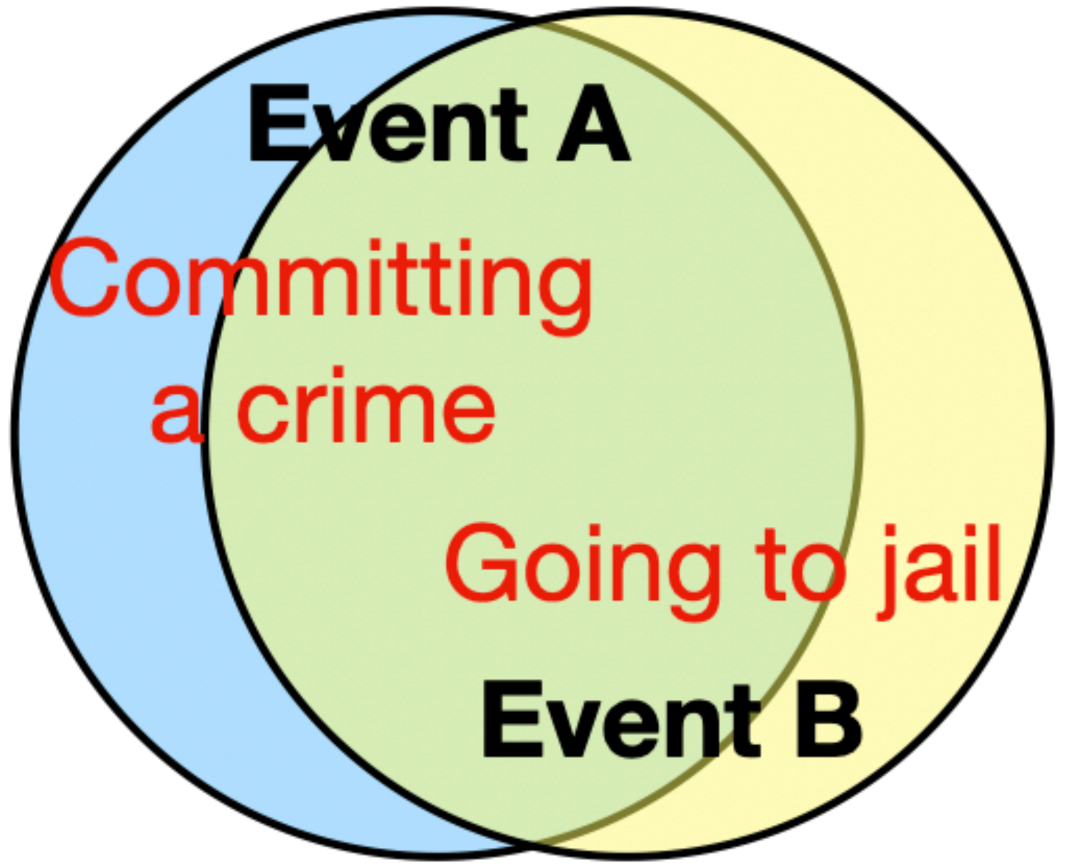
Dependent



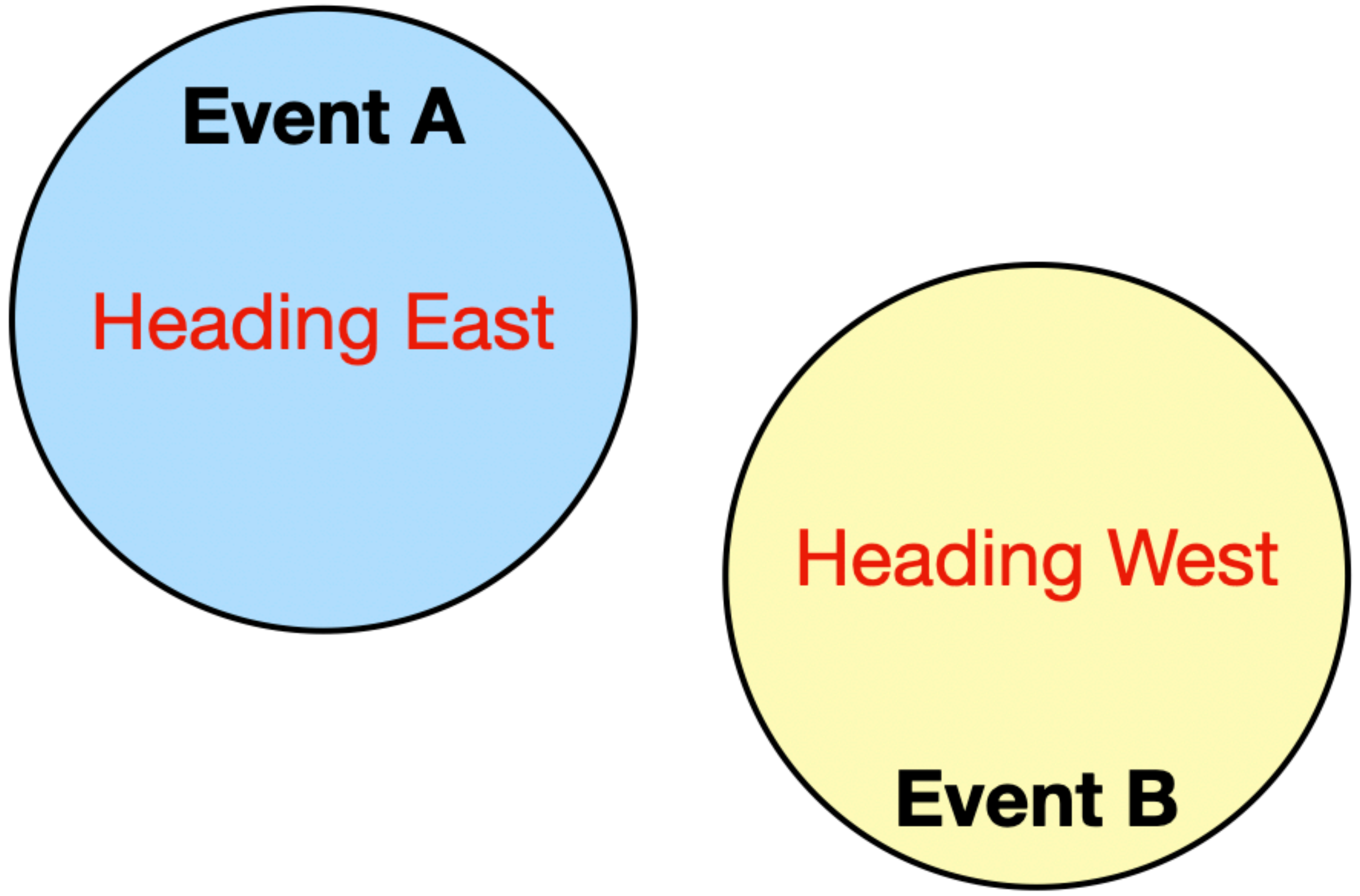
Defining **independent** events

- Two events are **independent** if the fact that one event has occurred **does not affect** the probability that the other event will occur

- Two events are **dependent** if the fact that one event has occurred **does affect** the probability that the other event will occur



Dependent

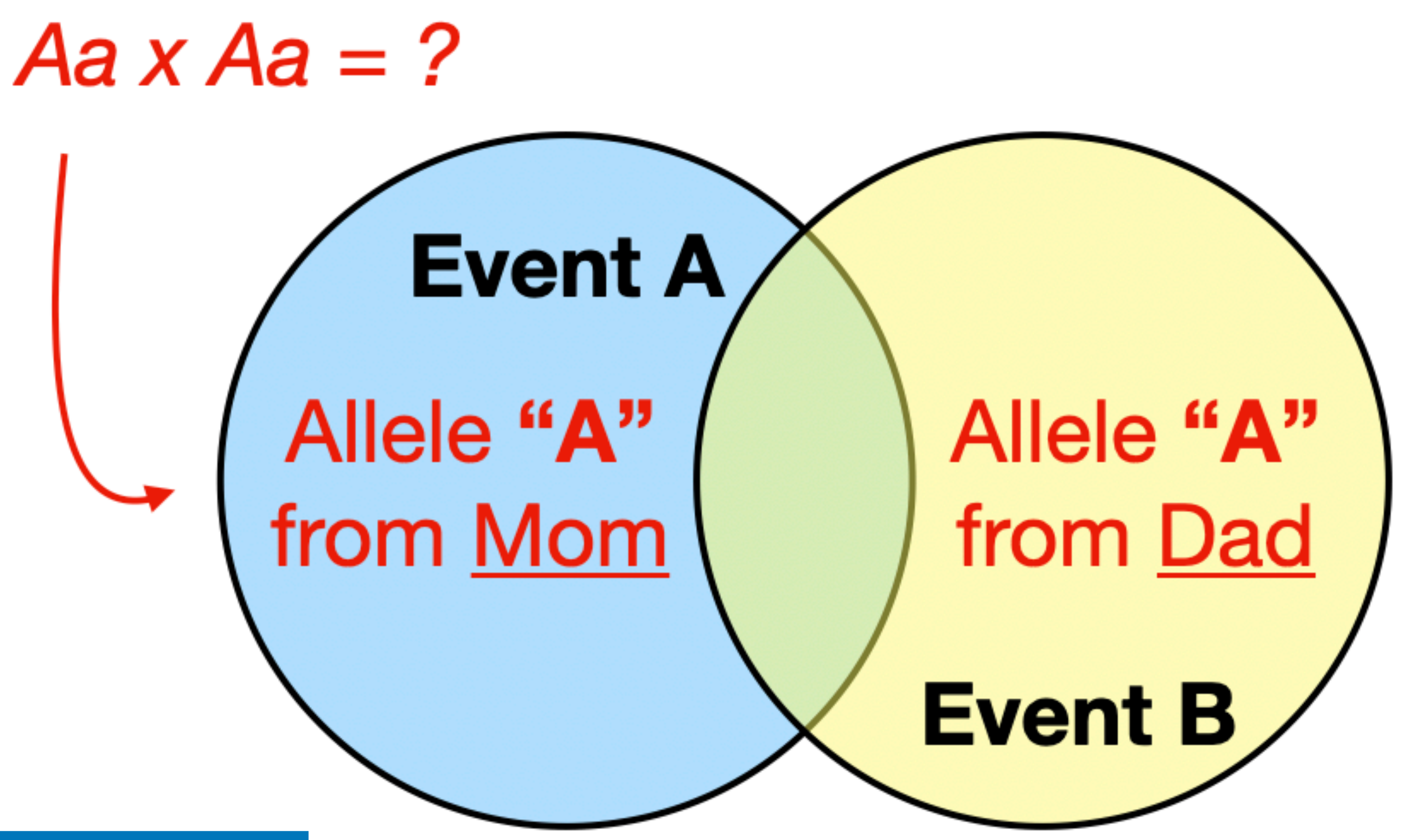


Dependent

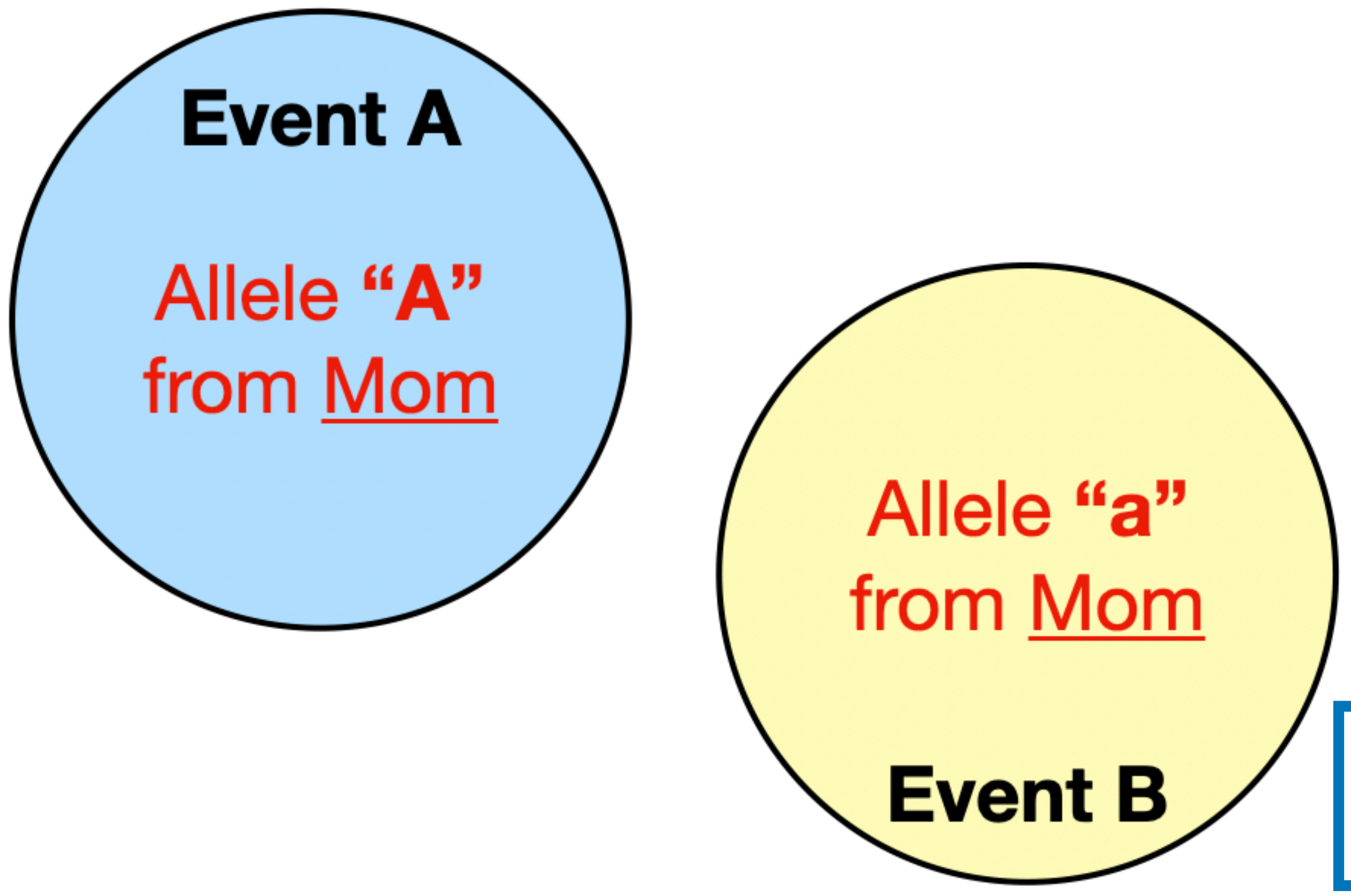
Defining **independent** events

- Two events are **independent** if the fact that one event has occurred **does not affect** the probability that the other event will occur

- Two events are **dependent** if the fact that one event has occurred **does affect** the probability that the other event will occur



Independent



Dependent

Defining **independent** and **disjoint** events

- If two events are **disjoint** they are, by default, **dependent**

	Independent	**Dependent**
Disjoint		Heading East; Heading West
Non-disjoint	Coin #1: Heads Coin #2: Heads	Committing crime; Going to jail

Rules of probability (part 2)

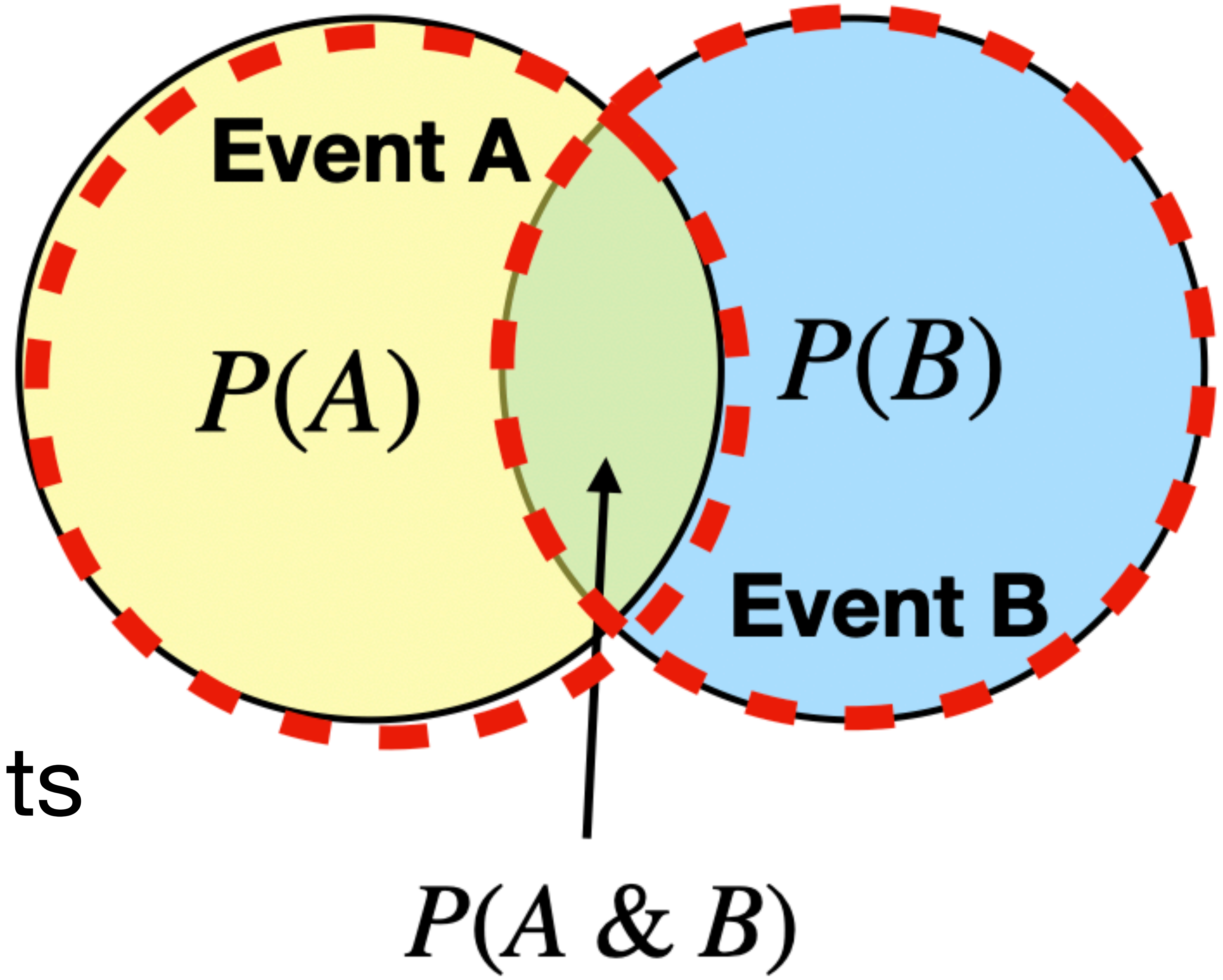
- **<u>Pr{A or B} = Pr{A} + Pr{B} - Pr{A and B}</u>**

    Pr{A and B} = 0 for all *disjoint* events

- **<u>Pr{A and B} = Pr{A} x Pr{B | A}</u>**

    Pr{B | A} = Pr{B} for all *independent* events

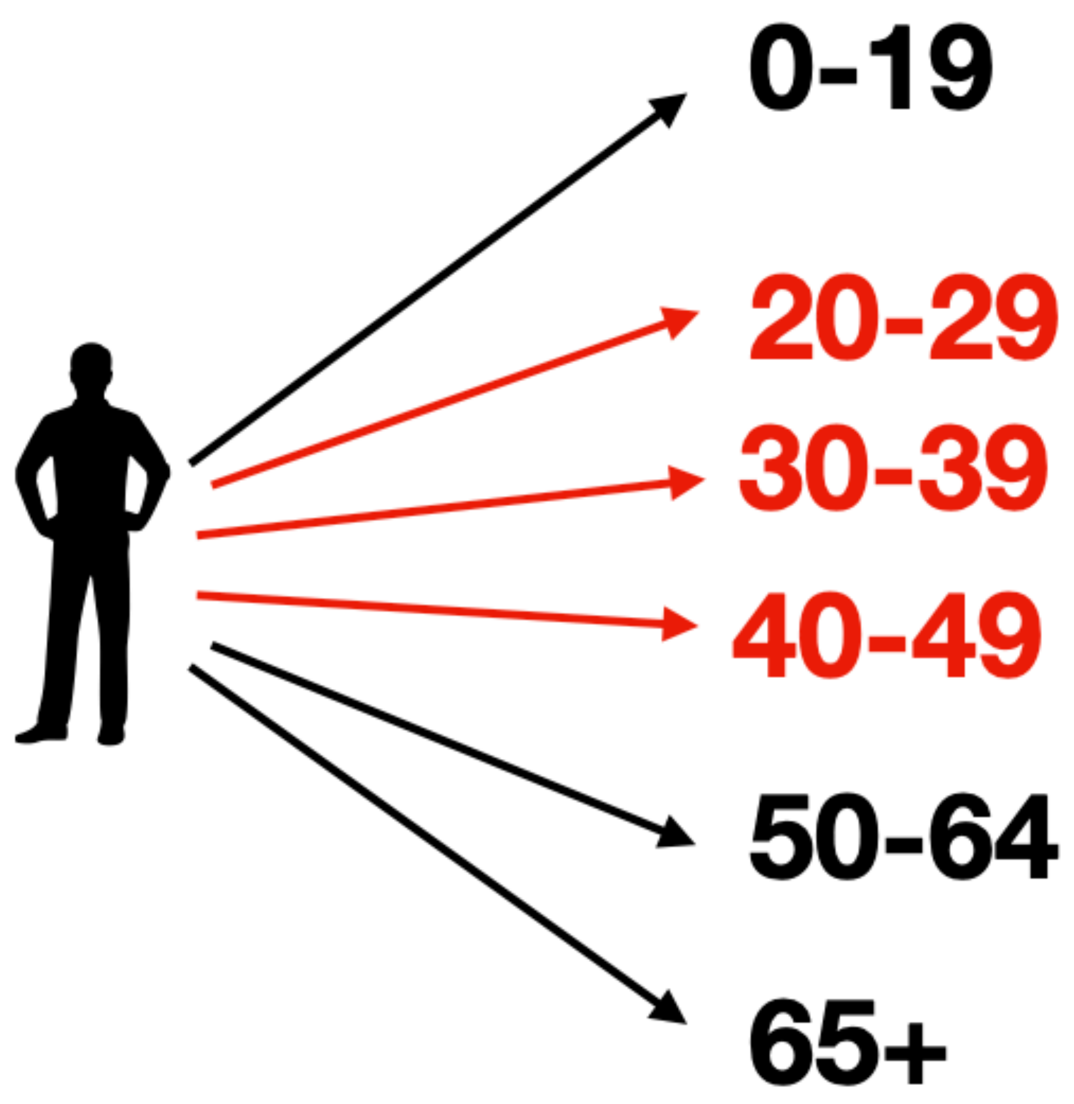
("B given A") = Conditional probability



Rules of probability (part 2)

- **Pr{A or B} = Pr{A} + Pr{B} - Pr{A and B}**

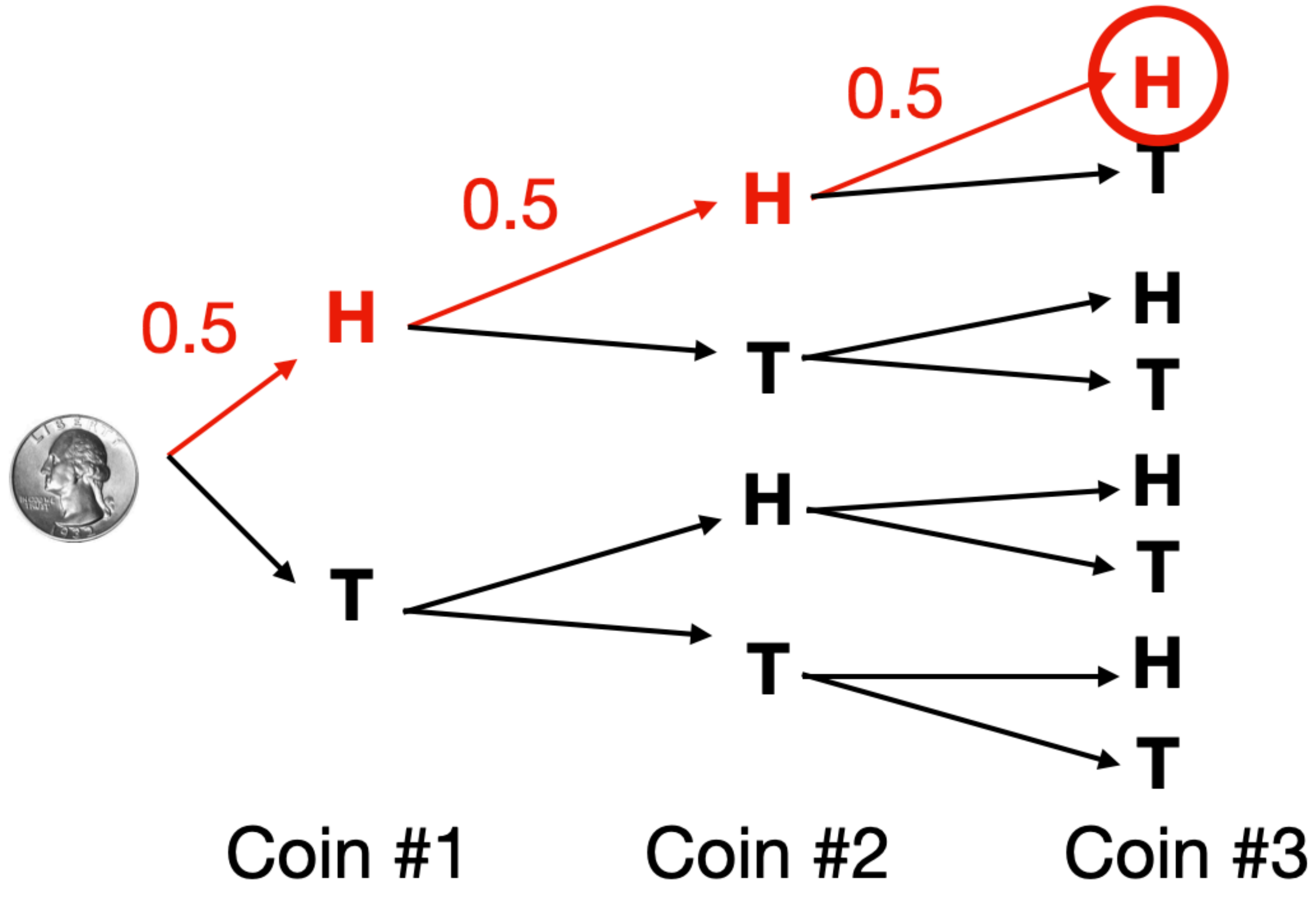
  Pr{A and B} = 0 for all *disjoint* events



- **Pr{A and B} = Pr{A} x Pr{B | A}**

  Pr{B | A} = Pr{B} for all *independent* events

("B given A") = Conditional probability



According to a recent study, 44% of the U.S. population has type O blood. Additionally, 15% of the population is Rh negative (and Rh status is independent of blood group). Thus, if someone is chosen at random, what is the probability that the person has type O, Rh negative blood?

According to a recent study, 44% of the U.S. population has type O blood. Additionally, 15% of the population is Rh negative (and Rh status is independent of blood group). Thus, if someone is chosen at random, what is the probability that the person has type O, Rh negative blood?

- *Independent events*
- *Non-disjoint events*
- *AND*

$$\Pr\{\text{group O \& Rh-}\} = \Pr\{\text{group O}\} \times \Pr\{\text{Rh-}\}$$

$$= 0.44 \times 0.15$$

$$= 0.066$$

What is the probability of randomly selecting a card from a standard 52-card deck that is either a heart or a spade?

(In a standard deck, 13 cards are hearts and 13 cards are spades)

What is the probability of randomly selecting a card from a standard 52-card deck that is either a heart or a spade?

(In a standard deck, 13 cards are hearts and 13 cards are spades)

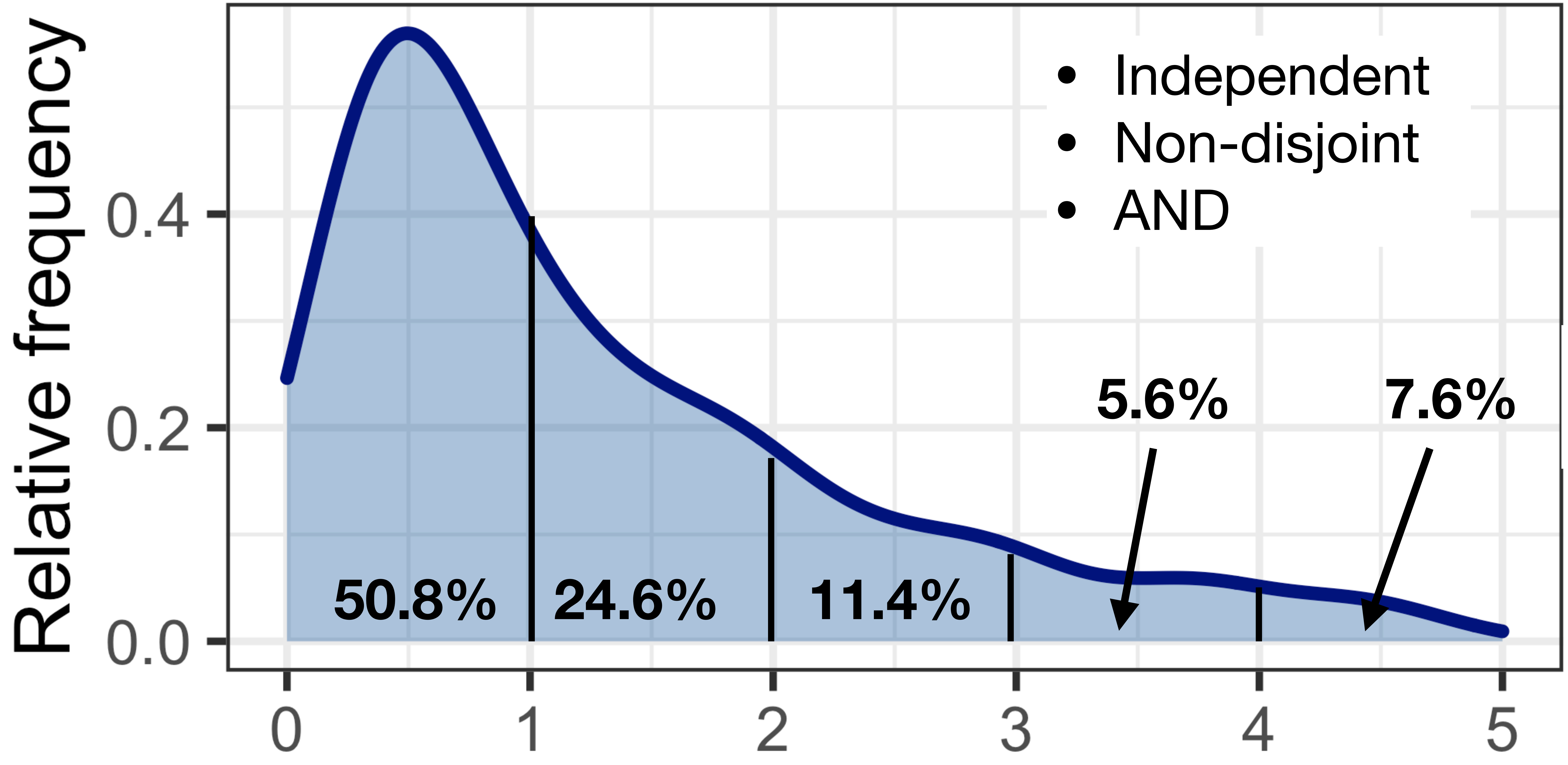
- *Disjoint events*
- *OR*

Pr{hearts OR spades} = Pr{hearts} + Pr{spades} - ~~Pr{hearts AND spades}~~

= 13/52 + 13/52 - 0

= 26/52 = 0.5

Density curves are used to estimate probabilities for continuous variables



Remember: probability = relative frequency (at high n)

Suppose you take a sample of two individuals, what is the probability that both values will be > 1?

Pr{Both > 1} = Pr{First > 1} * Pr{Second > 1}

Pr{value > 1} = 1 - 0.508

Pr{Both > 1} = (0.492)(0.492)

Pr{Both > 1} = 24.2%

	Income			
	Low	Medium	High	Total
Stressed	526	274	216	**1,016**
Not stressed	1,954	1,680	1,899	**5,533**
Total	**2,480**	**1,954**	**2,115**	**6,549**

What is the probability that someone in this study is stressed?

	Income			
	Low	Medium	High	Total
Stressed	526	274	216	**1,016**
Not stressed	1,954	1,680	1,899	**5,533**
Total	**2,480**	**1,954**	**2,115**	**6,549**

What is the probability that someone in this study is stressed?

Probability ~= relative frequency

$$\Pr\{\text{Stressed}\} = \frac{\#\ \text{Stressed}}{\#\ \text{Total}} = \frac{1{,}106}{6{,}549} = 0.1551 = 15.51\%$$

	Income			
	Low	Medium	High	Total
Stressed	526	274	216	**1,016**
Not stressed	1,954	1,680	1,899	**5,533**
Total	**2,480**	**1,954**	**2,115**	**6,549**

Given that someone in this study is from the highest income group, what is the probability that the person is stressed?

	Income			
	Low	Medium	High	Total
Stressed	526	274	216	**1,016**
Not stressed	1,954	1,680	1,899	**5,533**
Total	**2,480**	**1,954**	**2,115**	**6,549**

Given that someone in this study is from the highest income group, what is the probability that the person is stressed?

$$\Pr\{\text{Stressed} \mid \text{High income}\} = \frac{\#\ \text{Stressed (High income)}}{\#\ \text{Total (High income)}} = \frac{216}{2{,}115} = 10.21\%$$

	Income			
	Low	**Medium**	**High**	**Total**
Stressed	526	274	216	**1,016**
Not stressed	1,954	1,680	1,899	**5,533**
Total	**2,480**	**1,954**	**2,115**	**6,549**

Given that someone in this study is from the highest income group, what is the probability that the person is stressed?

$$\Pr\{\text{Stressed} \mid \text{High income}\} = \frac{\#\ \text{Stressed (High income)}}{\#\ \text{Total (High income)}} = \frac{216}{2{,}115} = 10.21\%$$

$$\boxed{\mathit{Pr\{B \mid A\} = Pr\{A \text{ and } B\} \,/\, Pr\{A\}}}$$

	Income			
	Low	Medium	High	Total
Stressed	526	274	216	**1,016**
Not stressed	1,954	1,680	1,899	**5,533**
Total	**2,480**	**1,954**	**2,115**	**6,549**

Is being stressed independent of having high income?

	Income			
	Low	**Medium**	**High**	**Total**
Stressed	526	274	216	**1,016**
Not stressed	1,954	1,680	1,899	**5,533**
Total	**2,480**	**1,954**	**2,115**	**6,549**

Is being stressed independent of having high income? **NO!**

$$\Pr\{\text{Stressed}\} \;\; ? \;\; \Pr\{\text{Stressed} \mid \text{High income}\}$$

$$15.5\% \neq 10.21\%$$

**of course we would like to have a statistical test to show the confidence of this result…*