Module 2B: Probability

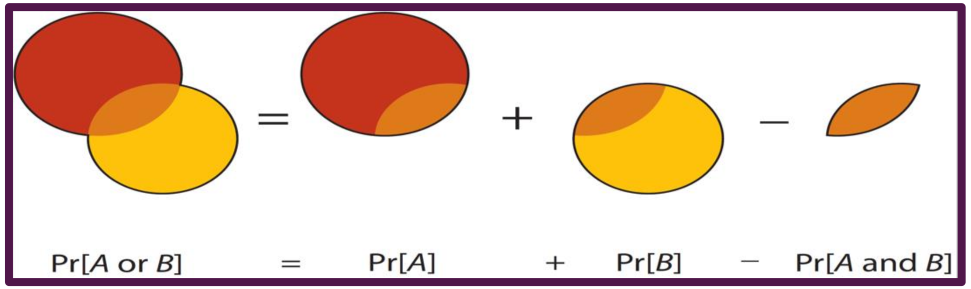
Frequentist and Bayesian building blocks

Agenda:

- **Frequentist probability**
  - Venn Diagram & definition of Event
  - **Addition Rule**
    - **Mutually exclusive, not mutually exclusive**
  - **Multiplication Rule**
    - **Independent, not independent**
  - Conditional Probability
  - Important differentiation: $P(A \cup B)$, $P(A \cap B)$, $P(A|B)$

Addition Rule:

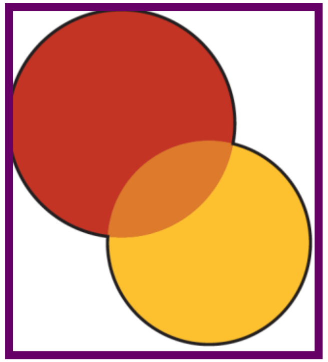
$$P[A\cup B] = P[A] + P[B] - P[A \cap B]$$

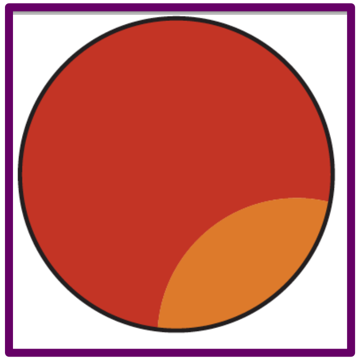


if mutually exclusive: $P[A\cup B]=P[A]+P[B]$

Multiplication Rule:

$$P[A\cap B]=P[A|B]P[B]=P[B|A]P[A]\xrightarrow{IFF independent}P[A]P[B]$$





- *Two events are independent if the occurrence of one gives no information about whether the second will occur*
- *Two events are dependent if the probability or outcome of one event <u>changes</u> because of the outcome of a second event*

<u>So far, we have learned a great many things about probability:</u>

1. <u>Sample space</u> is made up of elementary outcomes

2. <u>Events</u> can be elementary outcomes or groupings of elementary outcomes

3. Logic operators on probabilities: AND, NOT, OR

4. <u>General Addition rule</u>: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

5. IF events A and B are <u>mutually exclusive</u>, then general addition rule collapses into special addition rule: $P(A \cup B) = P(A) + P(B)$

6. <u>General Multiplication rule</u>: P[A and B]=P[A|B]xP[B] = P[B|A]XP[A]

7. If events A and B are independent, general multiplication rule collapses into special multiplication rule: P[A and B]=P[A]P[B]
   - *allows one to test whether or not two events are independent*

8. **What about if they are not independent?**

The Bruce Effect

A female house mouse mates and enters very early pregnancy. In her territory, **unfamiliar males** sometimes intrude. Exposure to the **scent of an unfamiliar male** (urine/pheromones) can trigger the **Bruce effect,** pregnancy block via implantation failure. This functions as an adaptive strategy in environments where infanticide risk from non-sire males is high: rather than invest in a likely doomed litter, the female aborts early and re-mates.

- If **no unfamiliar male is present**, pregnancy usually proceeds.
- If an **unfamiliar male is present**, pregnancy block is much more likely.

Thus, the **<u>state of the environment</u>** the female encounters (unfamiliar male present vs not) and the **<u>outcome</u>** (pregnancy continues vs blocks) are **dependent variables**

The Bruce Effect

State:

U(nfamiliar) =1 (present), =0 (absent)

Outcome:

B(lock) = 1, =0 (no block, pregnancy proceeds)

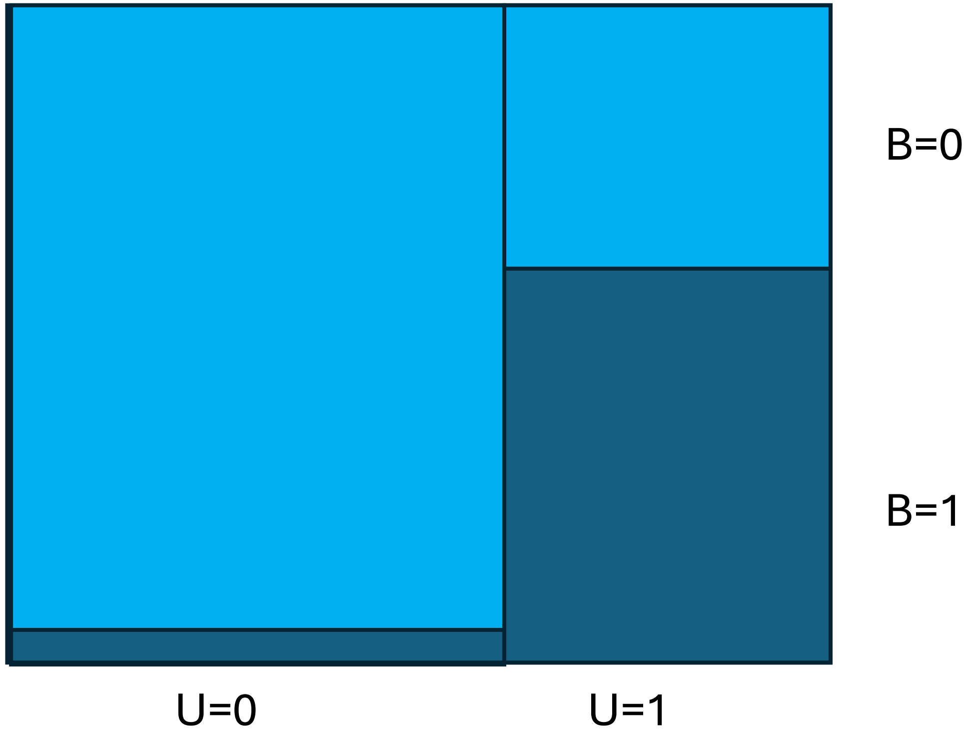
Prior:

P(U=1) = 0.40; P(U=0)=0.60

Conditional:

P(B=1|U=1) = 0.6

P(B=1|U=0)=0.05



These are made up numbers to make the calculations easier, but this paper characterized the Bruce effect!

https://www.nature.com/articles/184105a0

Another Example: *Nasonia vitripennis*, a parasitoid wasp, lays eggs in fly pupae; larval wasps then hatch inside, feed on host, and emerge as adults; the males and females then mate on the spot.

***Nasonia* females manipulate sex of their offspring depending on if host fly pupa previously parasitized**.

- If host not yet parasitized, then *Nasonia* lays mainly female eggs and produces only a few males (one male can fertilize multiple females).
- If host already parasitized, then *Nasonia* lays mostly male eggs.

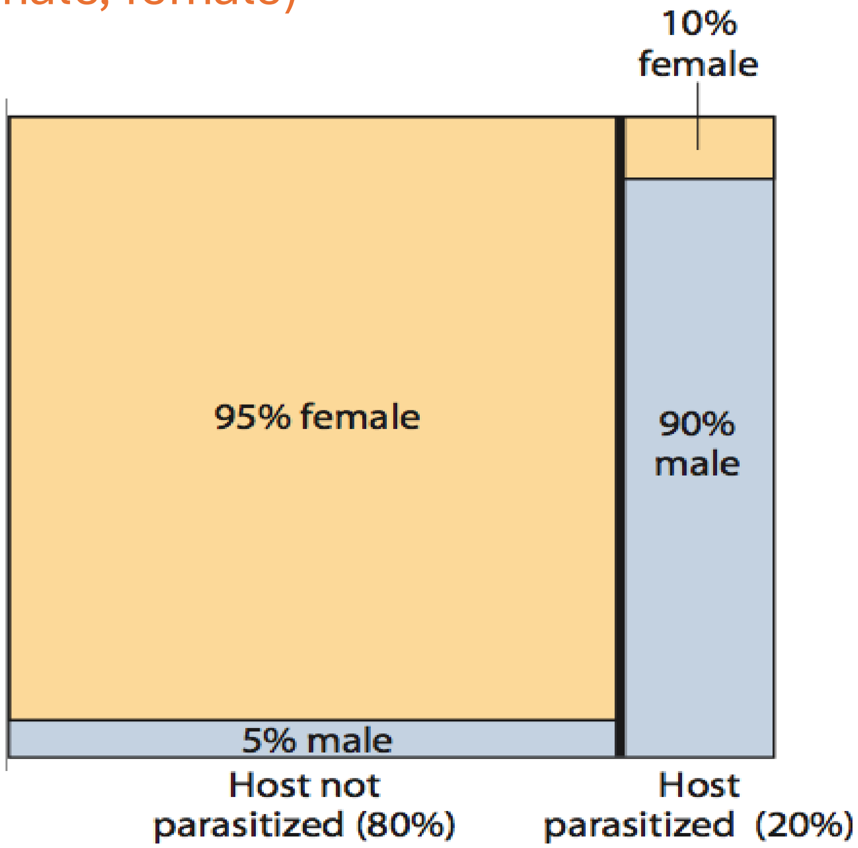
The state of the host encountered by a female and the sex of an egg laid are **dependent variables** (Werren, 1980)

- If host not yet parasitized, then *Nasonia* lays mainly female eggs and produces only a few males (one male can fertilize multiple females).
- If host already parasitized, then *Nasonia* lays mostly male eggs.

State of host (parasitized, not parasitized)

Possibly Dependent variable based on mosaic plot

Sex of egg (male, female)



Offspring of two carriers (Nn x Nn):

P[night blindness]=0.25

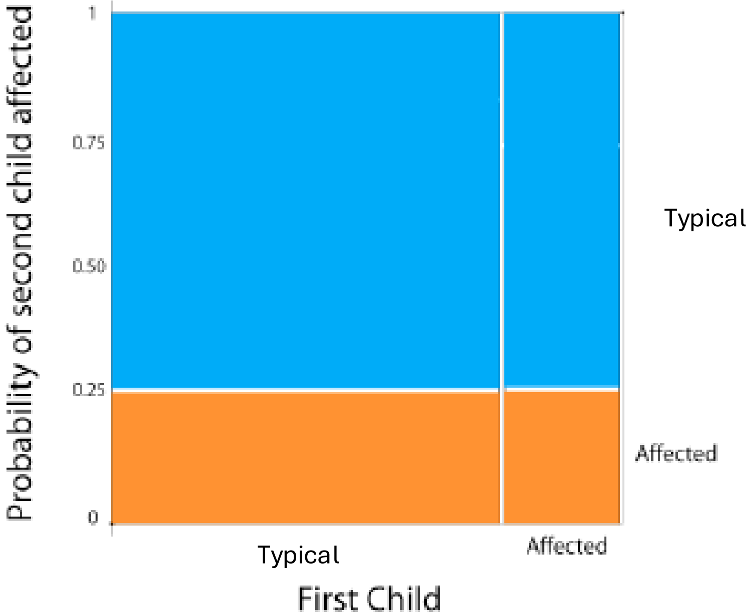
	N	**n**
N	NN	Nn
n	nN	nn

What is the probability that two kids from this family both have night blindness?

P[(1st child night blindness) AND (2nd child night blindness)]

= 0.25 x 0.25 = 0.0625

Possibly Independent variable based on mosaic plot:



Probability trees provide a straightforward method to determine independence or dependence between variables

- Map out probabilities of all mutually exclusive outcomes of variables

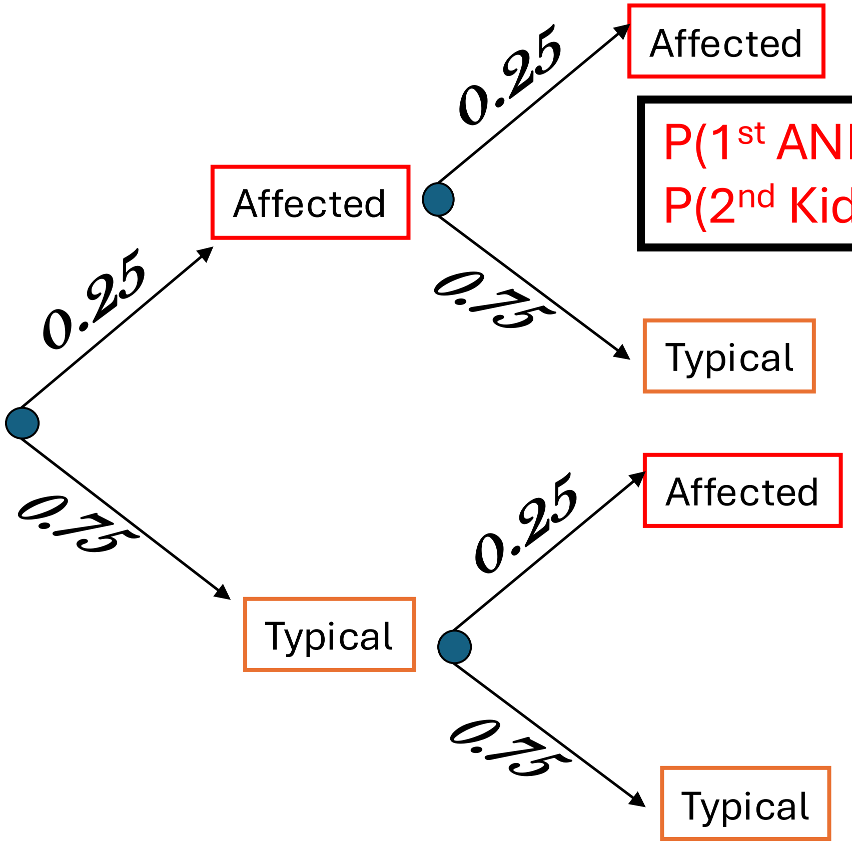
Additional Benefits:

- easy to calculate the probability of any possible outcome sequence for the variables under consideration

- easy to double check that all possibilities have been enumerated

Probability Trees: two events, meiosis is independent

Phenotype of 1^{st} Child | Phenotype of 2^{nd} Child | Probability:



$0.25 \times 0.25 = 0.0625$

P(1^{st} AND 2^{nd} kids affected) = P(2^{nd}|1^{st})P(1^{st})=
P(2^{nd} Kid affected) X P(1^{st} Kid affected)

$0.25 \times 0.75 = 0.1875$

$0.75 \times 0.25 = 0.1875$

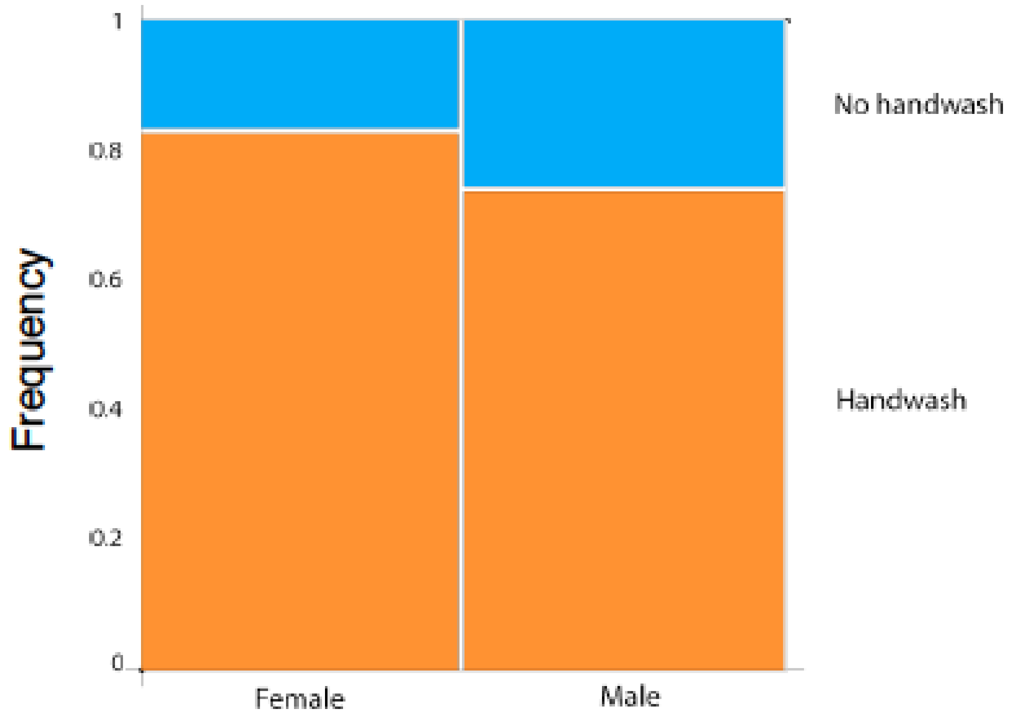
$0.75 \times 0.75 = 0.5625$

So far, we have learned a great many things about probability:

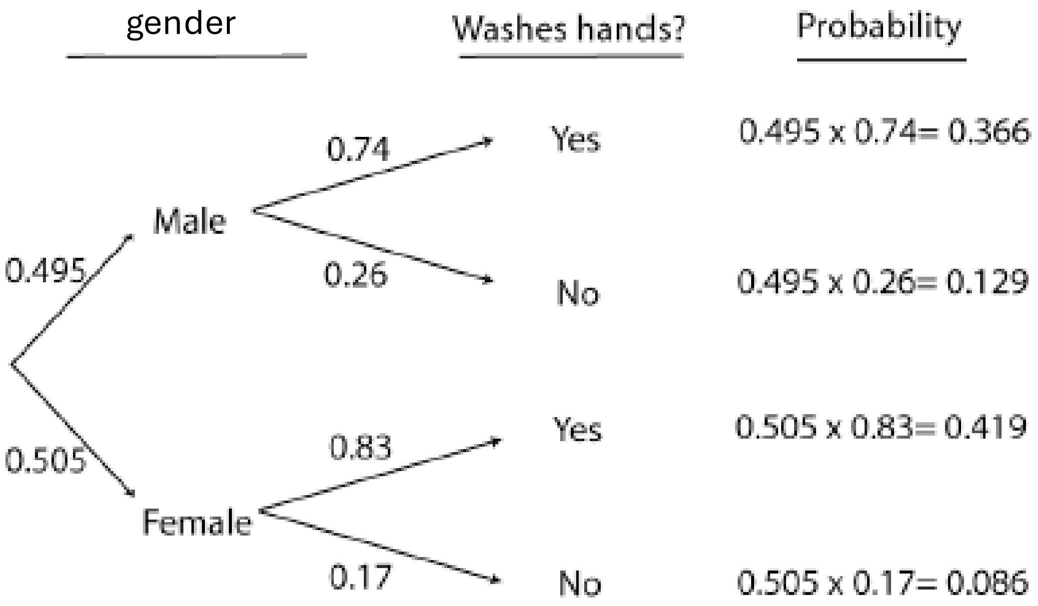
1. Sample space is made up of elementary outcomes
2. Events can be elementary outcomes or groupings of elementary outcomes
3. Logic operators on probabilities: AND, NOT, OR
4. General Addition rule: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
5. IF events A and B are mutually exclusive, then general addition rule collapses into special addition rule: $P(A \cup B) = P(A) + P(B)$
6. General Multiplication rule: P[A and B]=P[A|B]xP[B]
7. If events A and B are independent, general multiplication rule collapses into special multiplication rule: P[A and B] = P[A]xP[B]
    - allows one to test whether two events are independent

8. What about if they are not independent?

Example: Is washing your hands after using the washroom dependent on gender?



- P[male] = 0.495
- P[male washes his hands]=0.74
- P[female washes her hands]=0.83



These numbers were adapted from an earlier paper, but this is an actual phenomenon:
https://link.springer.com/article/10.1186/s12889-019-6705-5
(and, upsettingly, is a pattern among physicians, too)

Conditional Probability:

The probability that an event occurs given that a condition is met

P[X|Y] = P[X and Y]/P[Y]

This is read as “the probability of X given Y”

It means: the probability of X if Y is true

Fancier way of writing the **total probability** of an event:

$$\boxed{P[X] = \sum_{Y} P[X \mid Y] P[Y]}$$

Conditional Probability: **P[X|Y] = P[X and Y]/P[Y]**

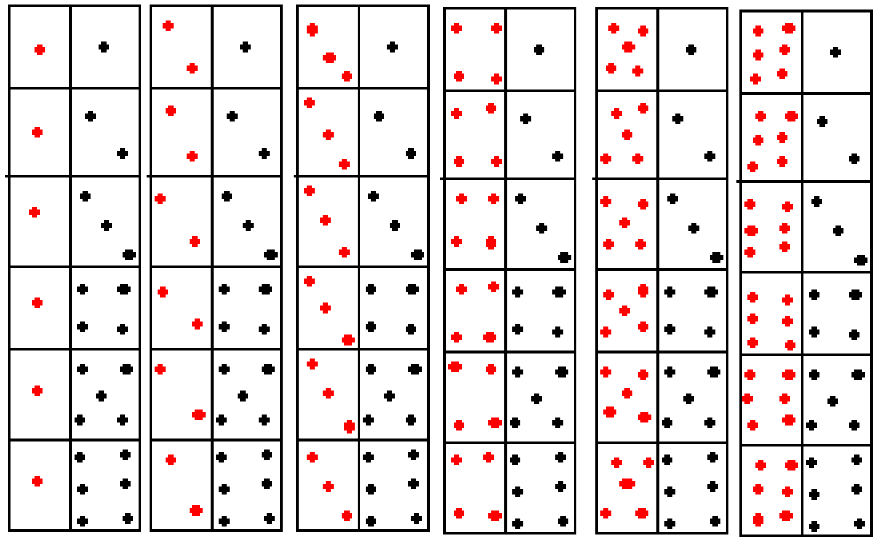
Example:

What is the probability that two dice will sum to three?

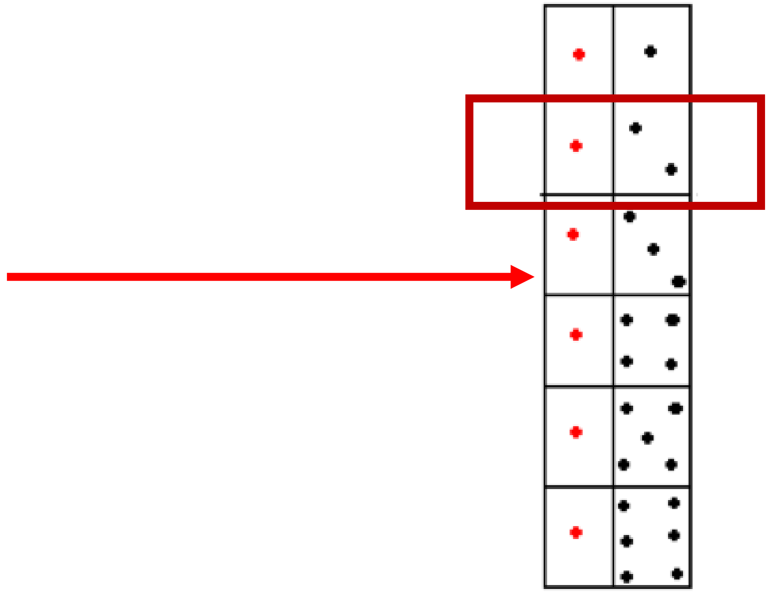
*-this is really asking **P[X and Y]** where $\boldsymbol{X}_{red\ die}$ =1 or 2 and $\boldsymbol{Y}_{black\ die}$ = 1 or 2*

Now: **what if we already have rolled the first die and know that we have a one? Event X_{red}=1**

Reduced state space, from 36 to 6:

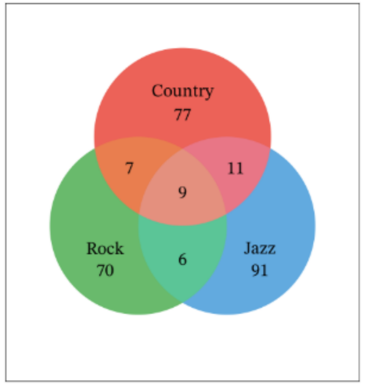


P[Sum to three] =2/36

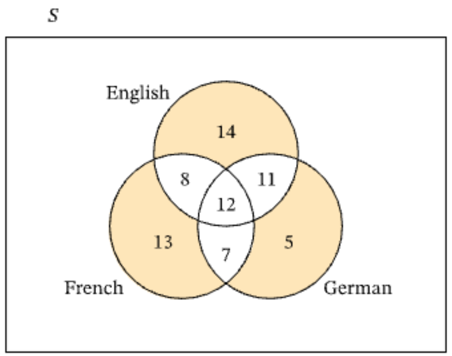


P[Sum to three]=1/6

- **Very** important to understand conditional probability before we tackle Bayes'
- Conditional probability can be a little confusing; sometimes using a Venn diagram with **3** events instead of 2 makes it clearer (example 5 and 6): https://www.nagwa.com/en/explainers/403141497934/



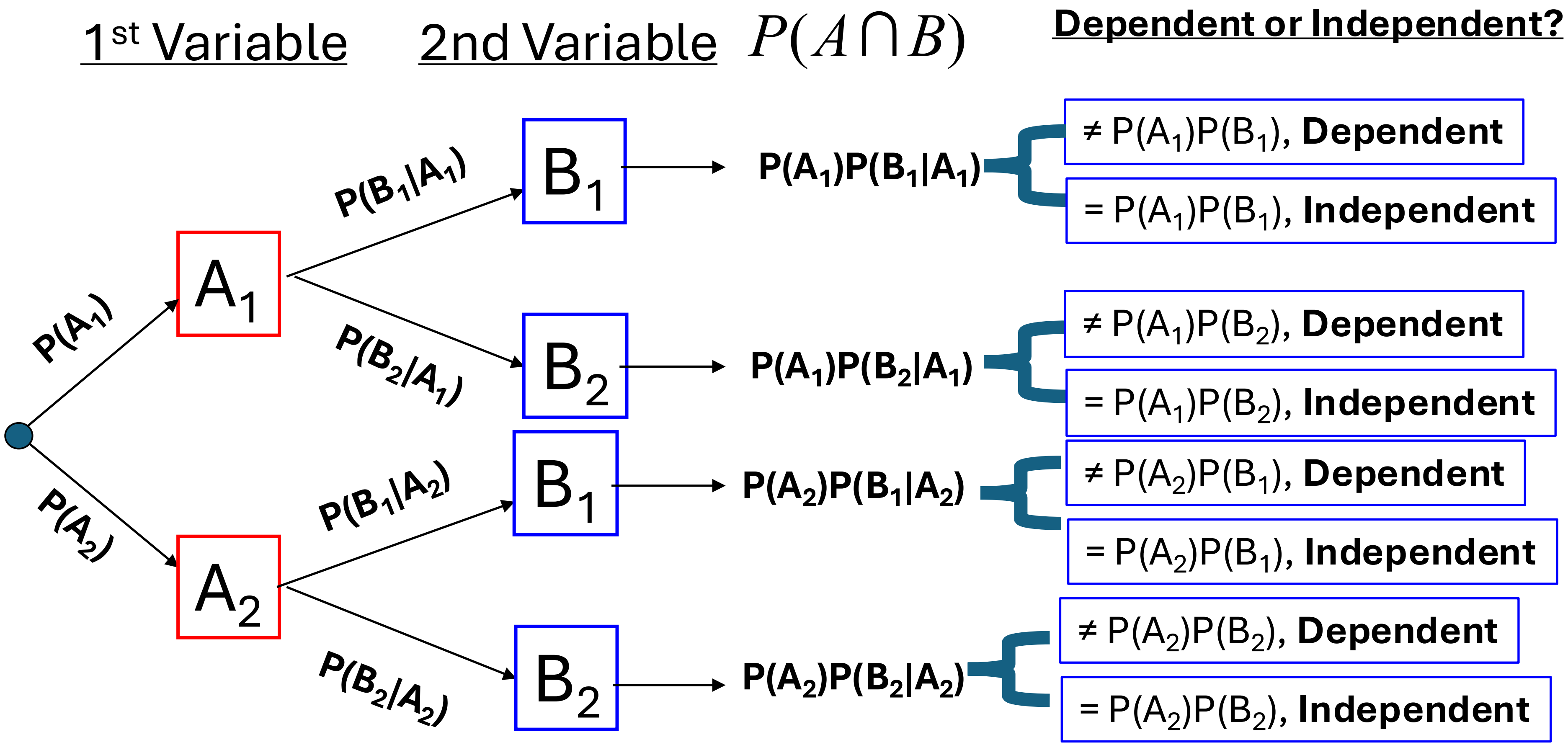
271 students voted for the types of music they wanted at the school dance. The results are shown in the Venn diagram. Find the probability that a randomly selected student voted for rock and not jazz.



In a sample of 100 students enrolling in a university, a questionnaire indicated that 45 of them studied English, 40 studied French, 35 studied German, 20 studied both English and French, 23 studied both English and German, 19 studied both French and German, and 12 studied all three languages.
Using a Venn diagram, find the probability that a randomly chosen student studied **only** one of the three languages.

- Some learners have struggled with the difference between $P(A \cap B)$ and P(A|B)

- **Remember:** $\boxed{P(A \cap B) = P(A)P(B \mid A) = P(B)P(A \mid B)}$


1st Variable
2nd Variable
$P(A \cap B)$
Dependent or Independent?
$P(A_1)$
$P(A_2)$
A_1
A_2
$P(B_1|A_1)$
$P(B_2|A_1)$
$P(B_1|A_2)$
$P(B_2|A_2)$
B_1
B_2
B_1
B_2
$P(A_1)P(B_1|A_1)$
$\neq P(A_1)P(B_1)$, Dependent
$= P(A_1)P(B_1)$, Independent
$P(A_1)P(B_2|A_1)$
$\neq P(A_1)P(B_2)$, Dependent
$= P(A_1)P(B_2)$, Independent
$P(A_2)P(B_1|A_2)$
$\neq P(A_2)P(B_1)$, Dependent
$= P(A_2)P(B_1)$, Independent
$P(A_2)P(B_2|A_2)$
$\neq P(A_2)P(B_2)$, Dependent
$= P(A_2)P(B_2)$, Independent