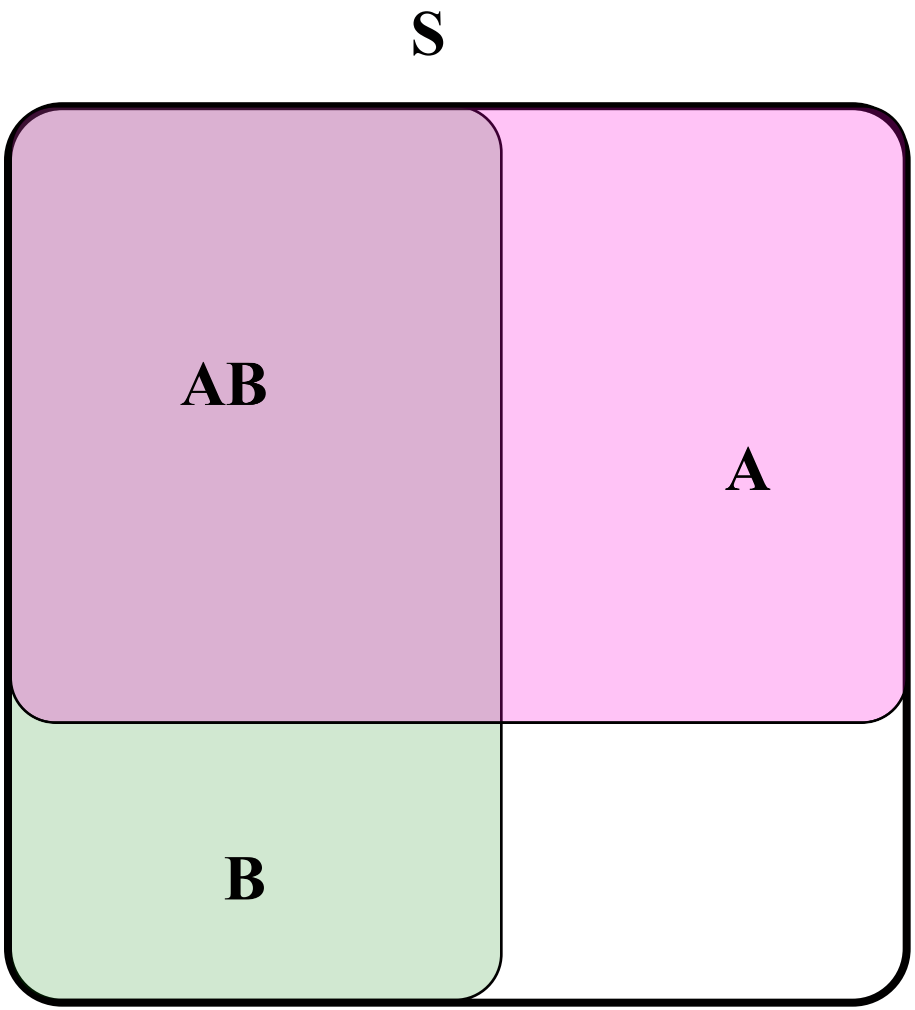
# What does independence look like?

# Independence




Independence Definition 1:

$$P(AB) = P(A)P(B)$$

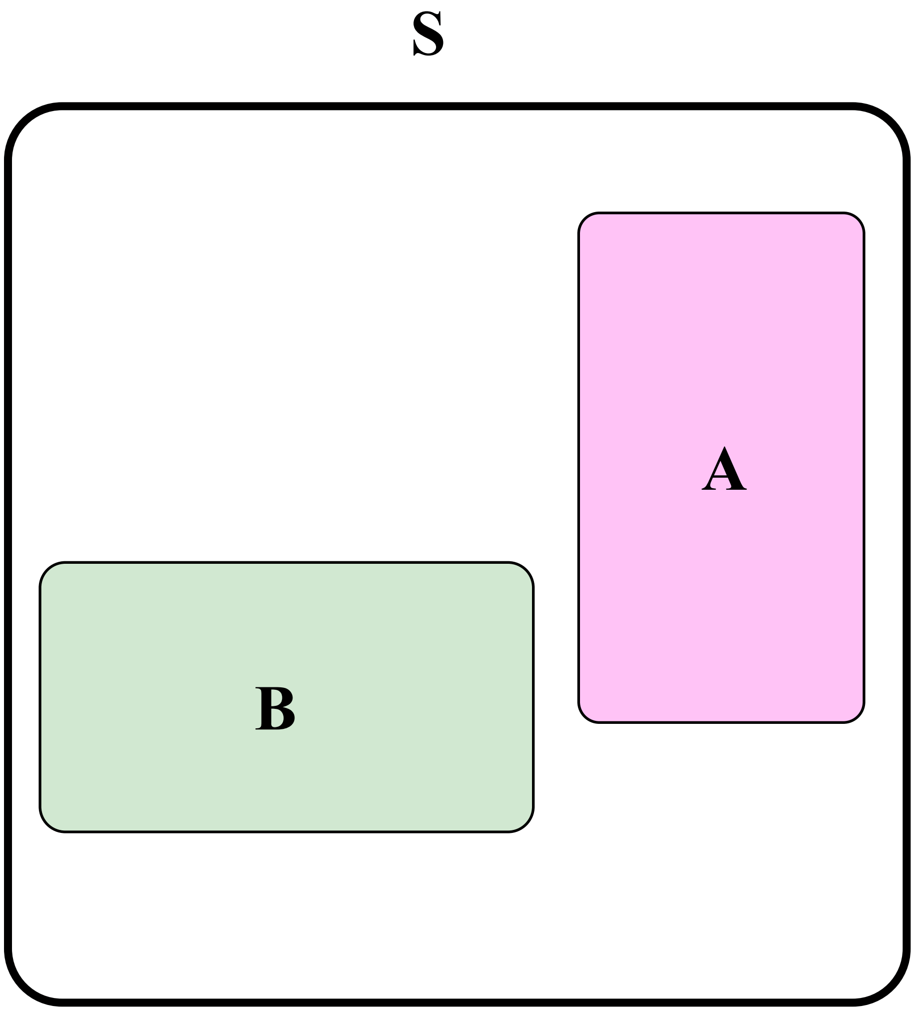
$$\frac{|AB|}{|S|} = \frac{|A|}{|S|} \times \frac{|B|}{|S|}$$

Independence Definition 2:

$$P(A|B) = P(A)$$

$$\frac{|AB|}{|B|} = \frac{|A|}{|S|}$$

# Independence?



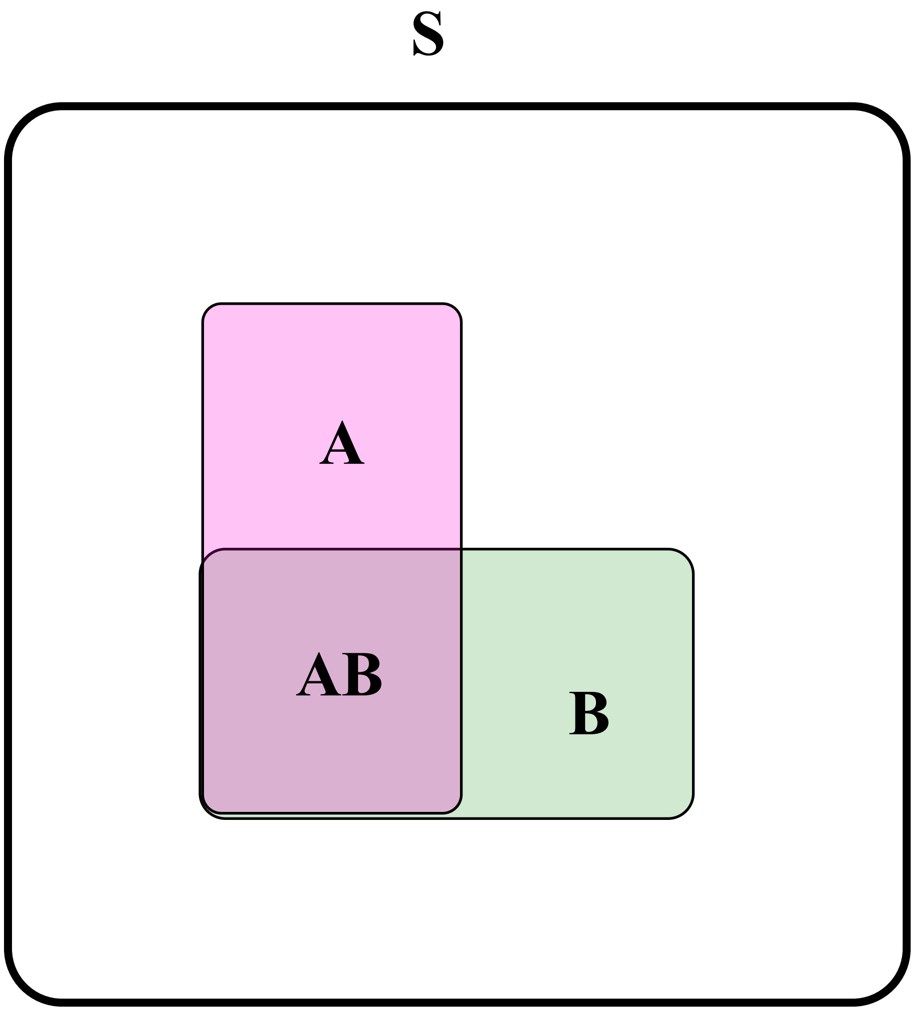

Independence Definition 1:

$$P(AB) = P(A)P(B)$$

$$\frac{|AB|}{|S|} = \frac{|A|}{|S|} \times \frac{|B|}{|S|}$$

0

# Independence?




Independence Definition 2:

$$P(A|B) \stackrel{?}{=} P(A)$$

$$\frac{|AB|}{|B|} \stackrel{?}{=} \frac{|A|}{|S|}$$

$$\frac{1}{2} \neq \frac{2}{16}$$

When we introduced conditions

Identities of probability remained the same

But sometimes independence / dependence relationships change

Current goals:

1) Recognize conditional independence / dependence.

2) Get Intuition

Future goal:

Use conditional independence / dependence in Machine Learning

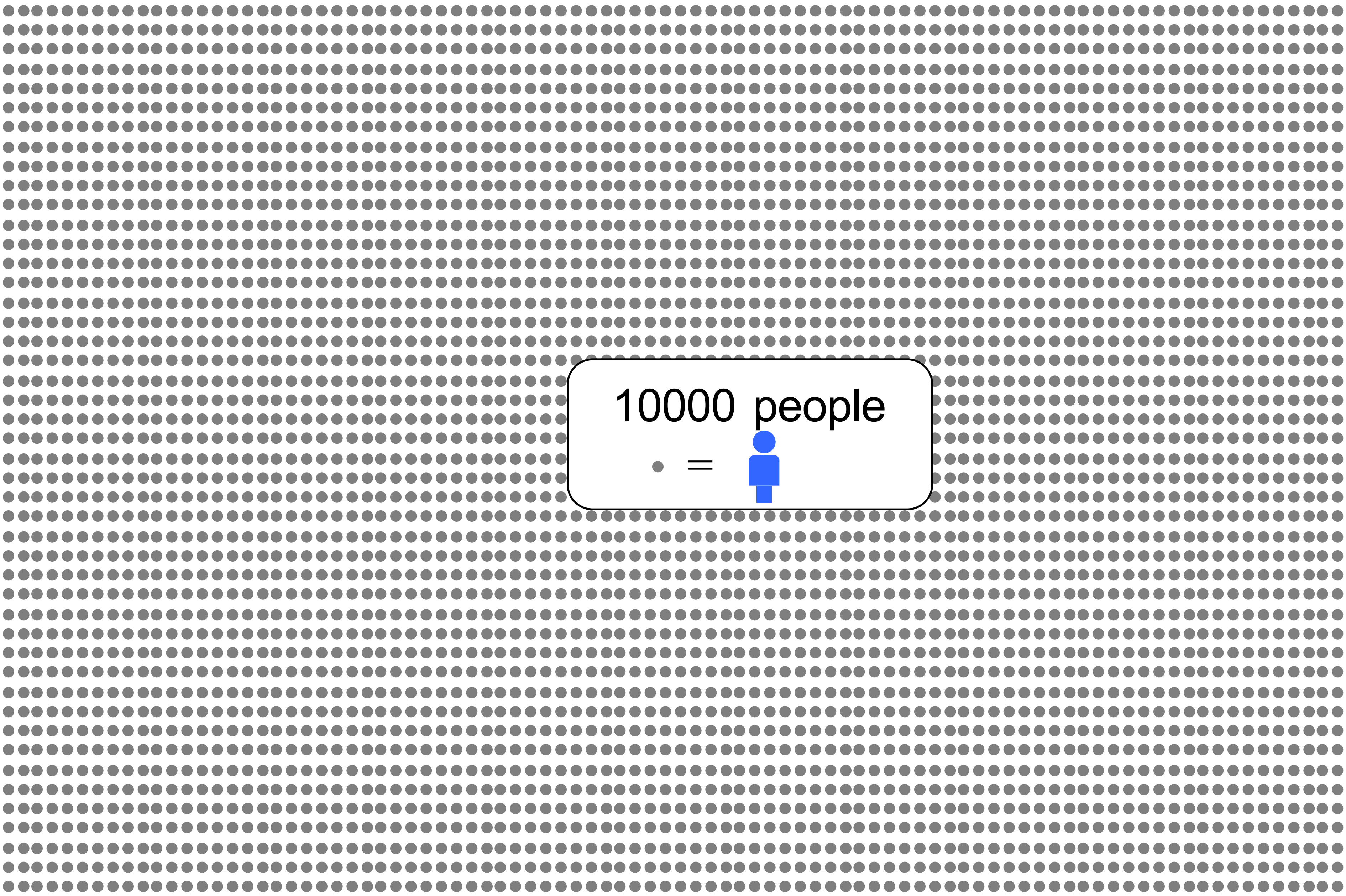
# Friday Night Fever

- Population of 10,000 people.
  - Of those, 300 have Malaria (event M) and 200 have Bacterial Infection (event B). 6 people have both.
  - Have Fever if and only if you have Malaria or Bacteria.
  - Are M and B independent?
- Solution:
  - P(M) = 300 / 10,000 = 0.03
  - P(B) = 200 / 10,000 = 0.02
  - P(MB) = 6 / 10,000 = 0.0006
  - P(M)P(B) = 0.0006
  - P(M)P(B) = P(MB)
  - Independent

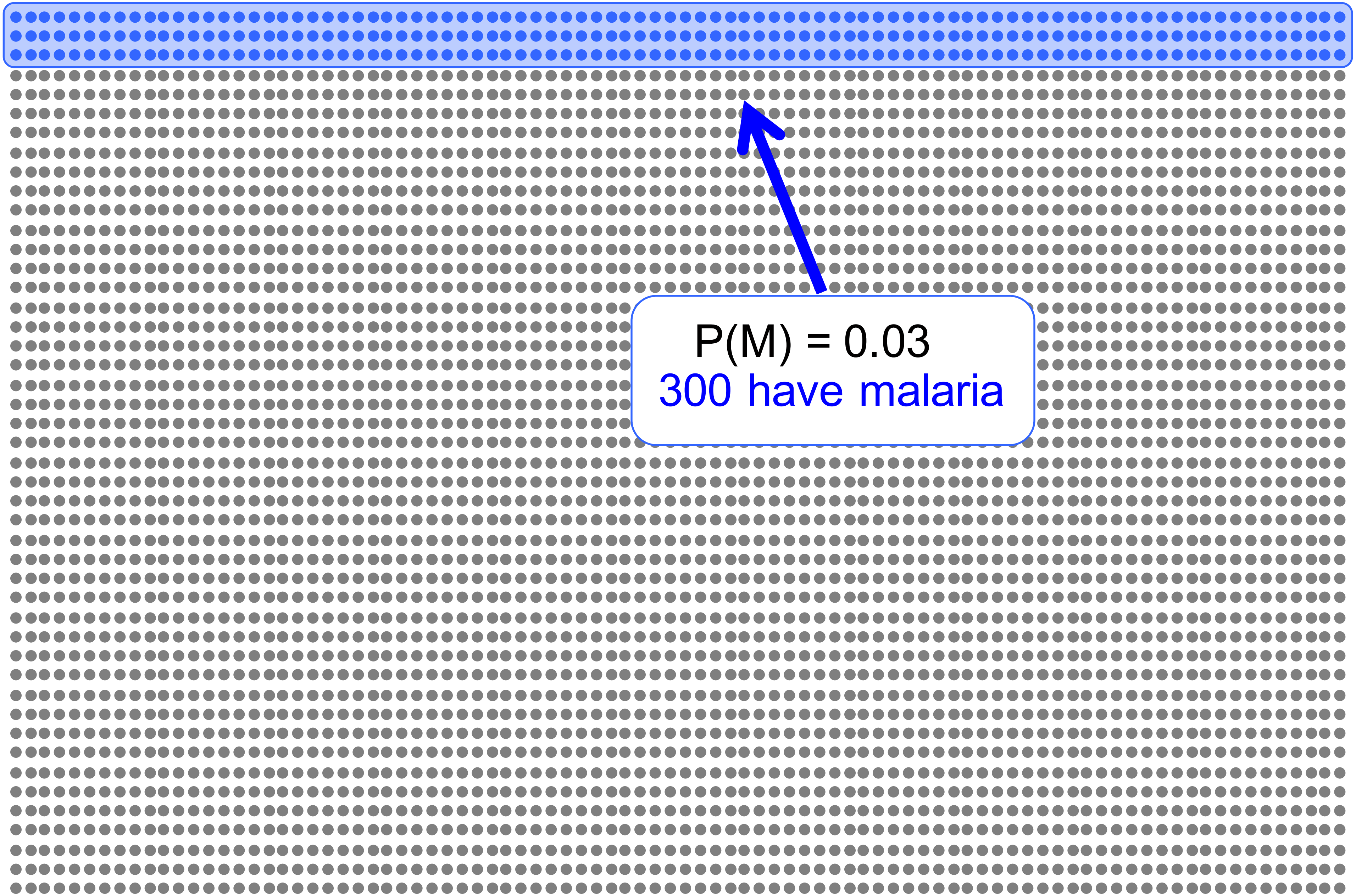
# Friday Night Fever

- Population of 10,000 people.
  - Of those, 300 have Malaria (event M) and 200 have Bacterial Infection (event B). 6 people have both.
  - Have Fever if and only if you have Malaria or Bacteria.
  - Are M and B independent given F?
- Solution:
  - Total people with Fever = 200+300 – 6 = 494
  - $P(M|F) = 300 / 494 = 0.61$
  - $P(B|F) = 200 / 494 = 0.40$
  - $P(MB|F) = 6 / 494 = 0.012$
  - $P(M|F)P(B|F) = 0.224$
  - $P(M|F)P(B|F) \neq P(MB|F)$
  - Conditionally dependent

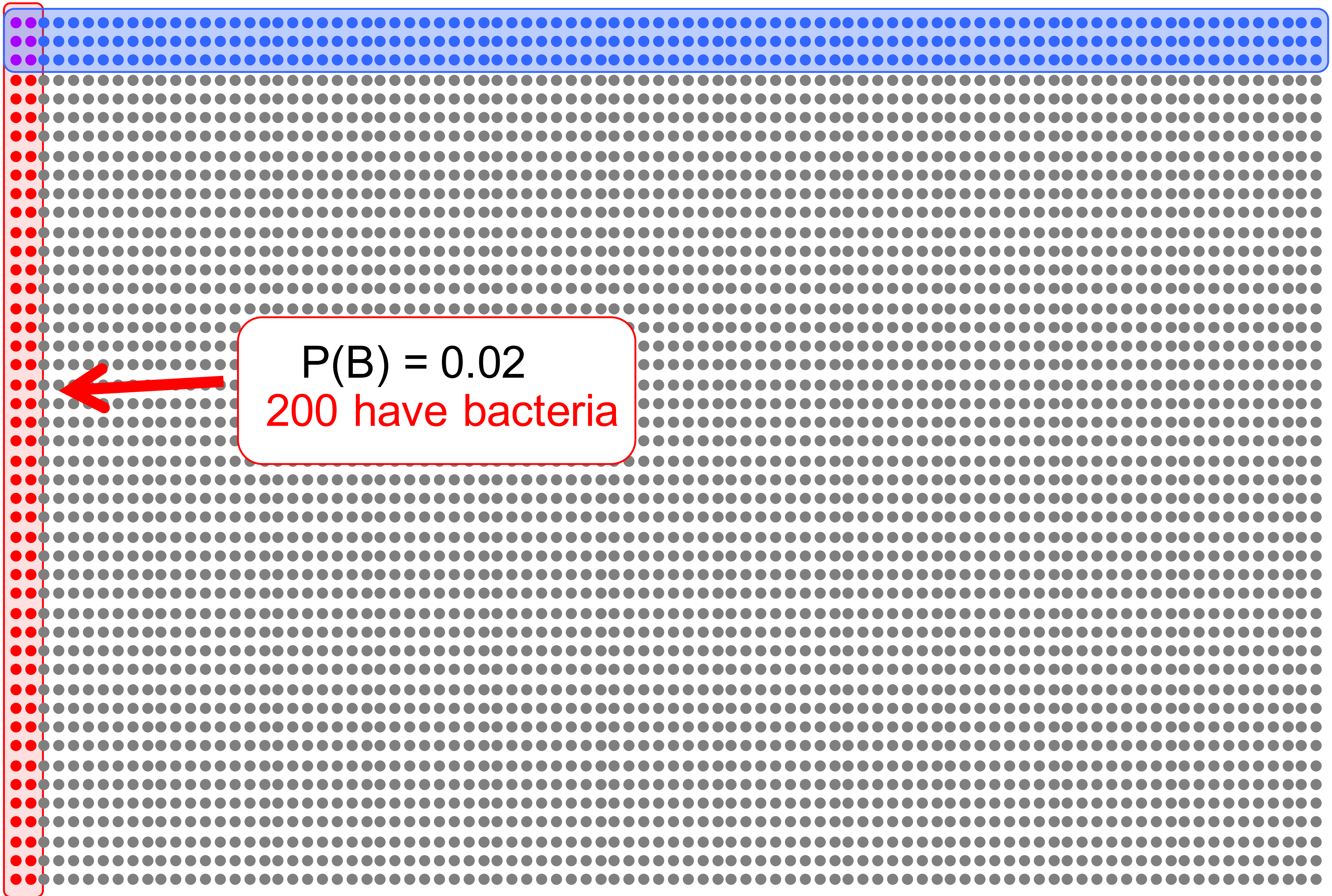
# Conditional Dependence



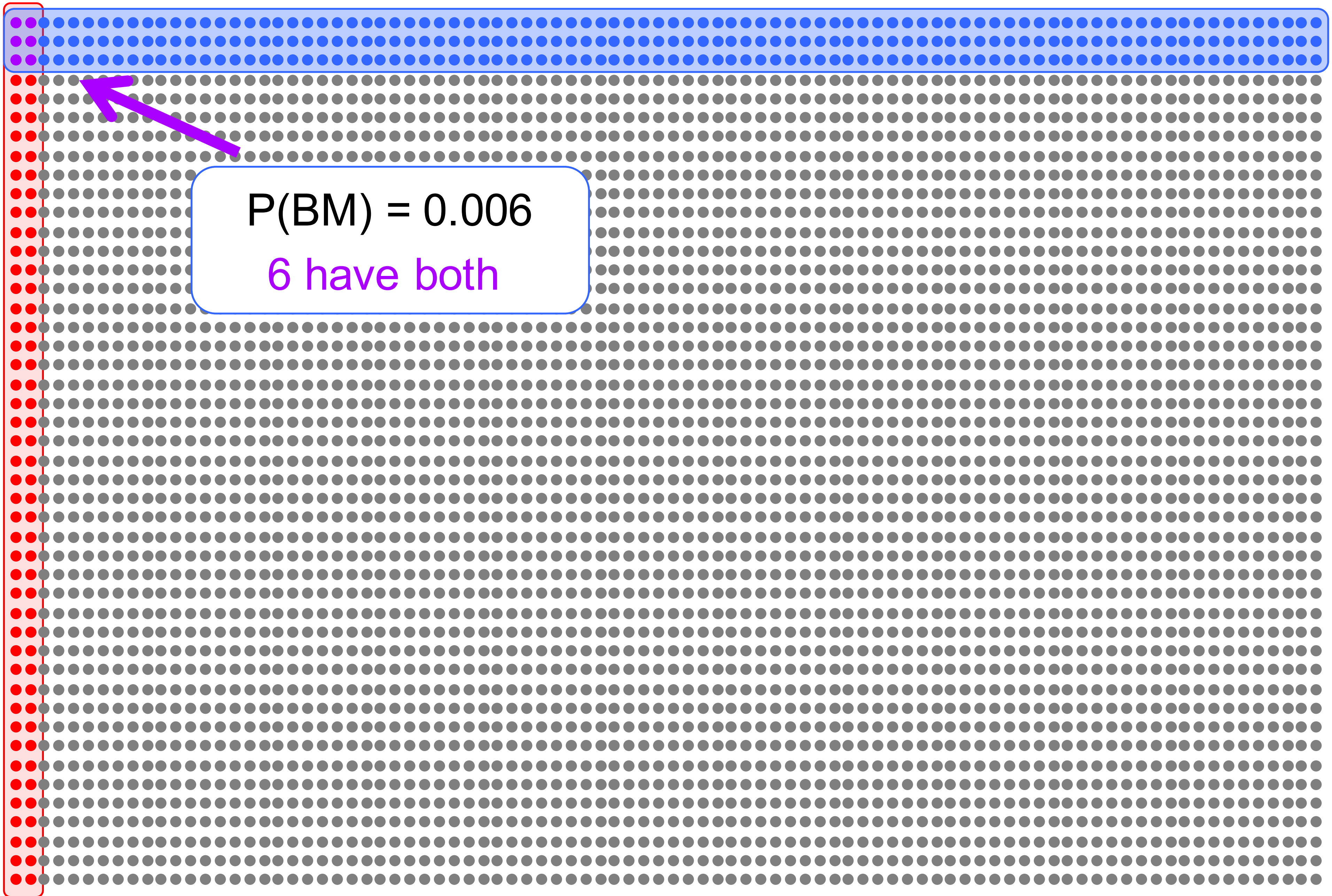
# Conditional Dependence



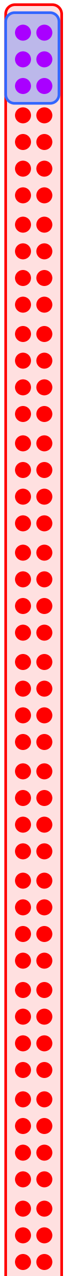
# Conditional Dependence



# Conditional Dependence



# Conditional Dependence

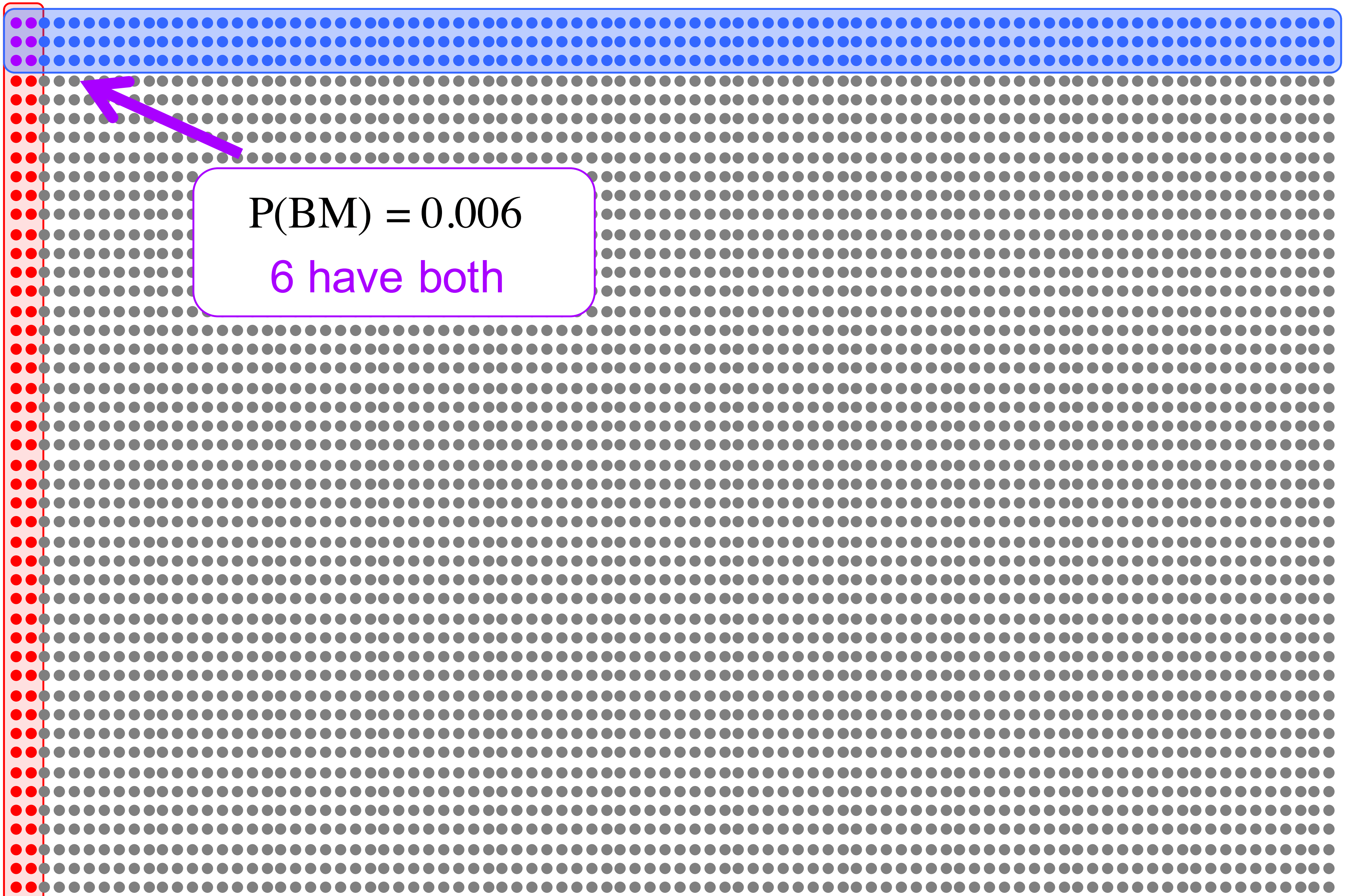


If we condition on B, the same ratio of people have malaria

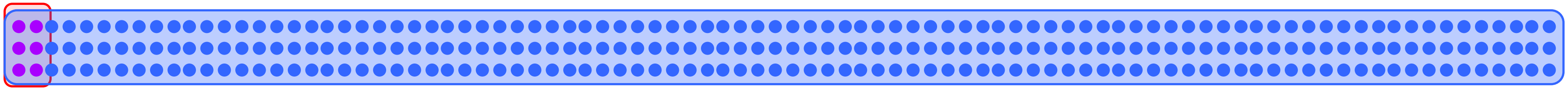
$$P(M|B) = 6/200 = 0.03$$
$$P(M) = 300/10000 = 0.03$$
$$P(M) = P(M|B)$$

That's the math definition of independence

# Conditional Dependence



# Conditional Dependence

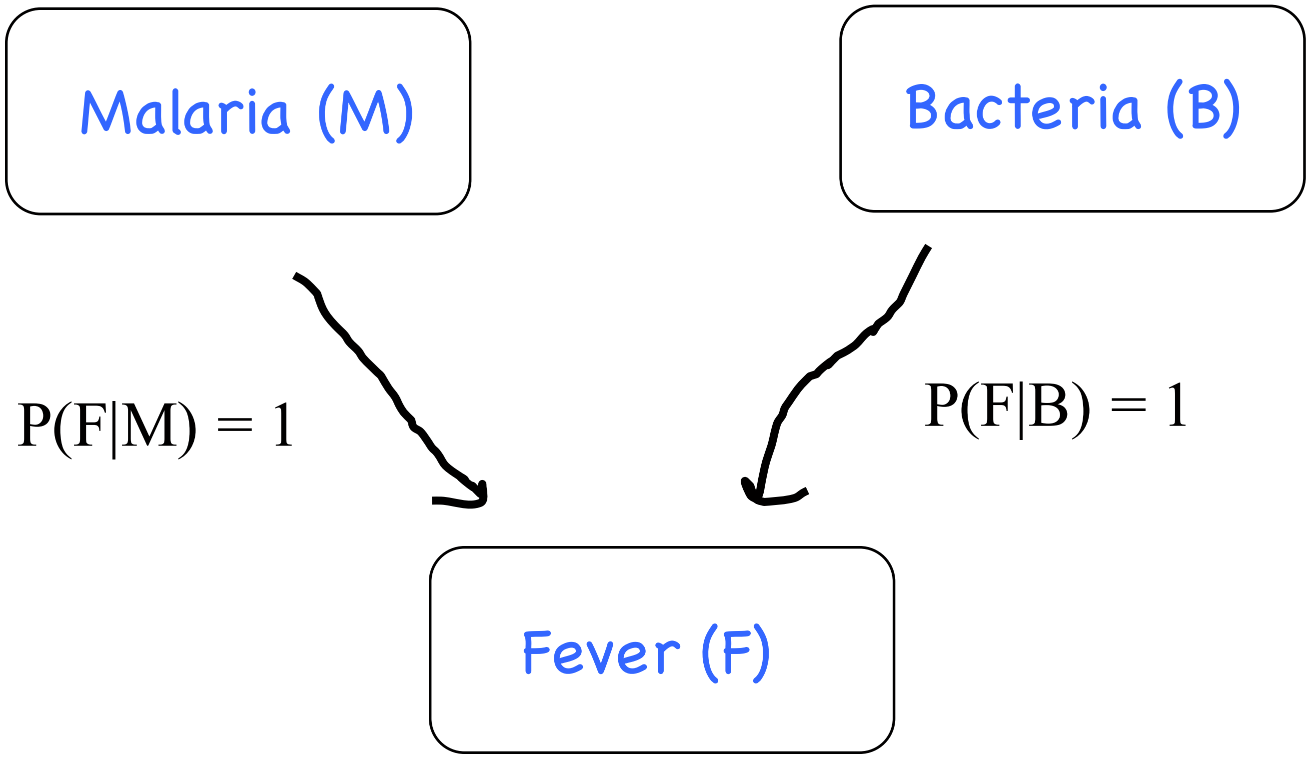


If we condition on M, the same ratio of people have bacteria

There it is again!

$$P(B|M) = 6/300 = 0.02$$
$$P(B) = 200/10000 = 0.02$$
$$P(B|M) = P(B)$$

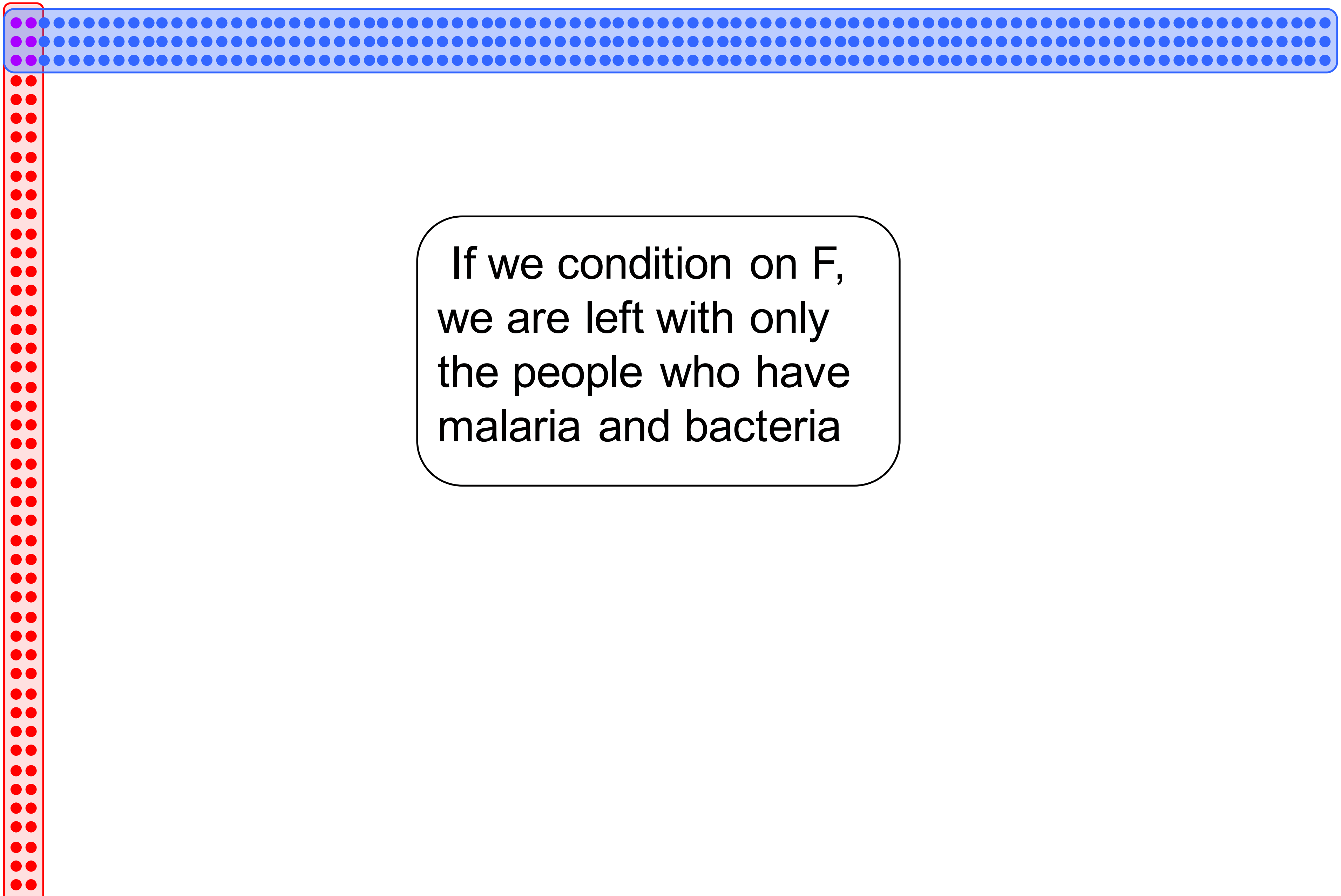
# Conditional Dependence




*This is a “causal” diagram. It helps explain why things are independent

# Conditional Dependence

# Conditioned on Fever

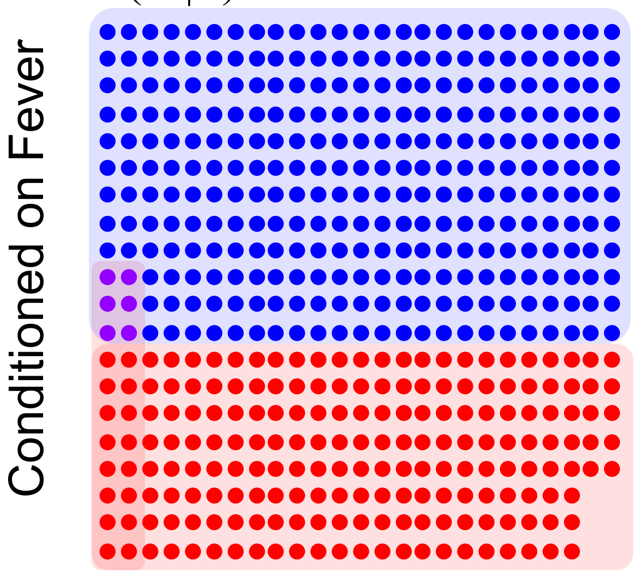


If we condition on F, we are left with only the people who have malaria and bacteria

# Conditioned on Fever

P(B|F) = 200/494 = 0.40

P(M|F) = 300/494 = 0.61

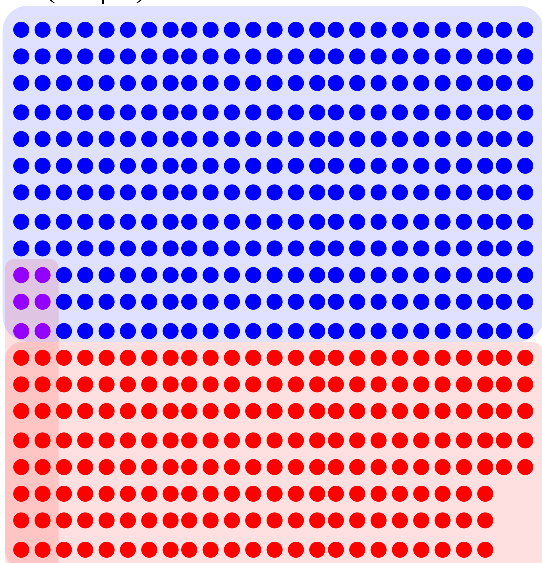


# Conditioned on Fever

$P(B|F) = 200/494 = 0.40$

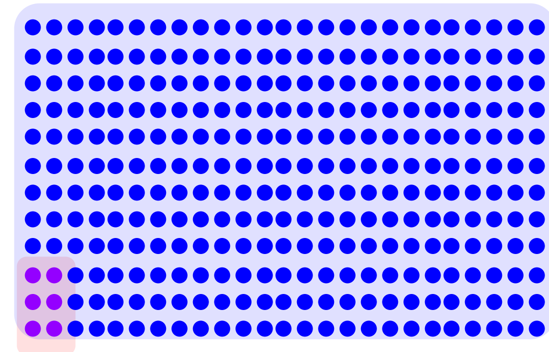
$P(M|F) = 300/494 = 0.61$

Conditioned on Fever



Test shows Malaria →

Conditioned on Fever + Malaria



$P(B|MF) = 6/300 = 0.02$

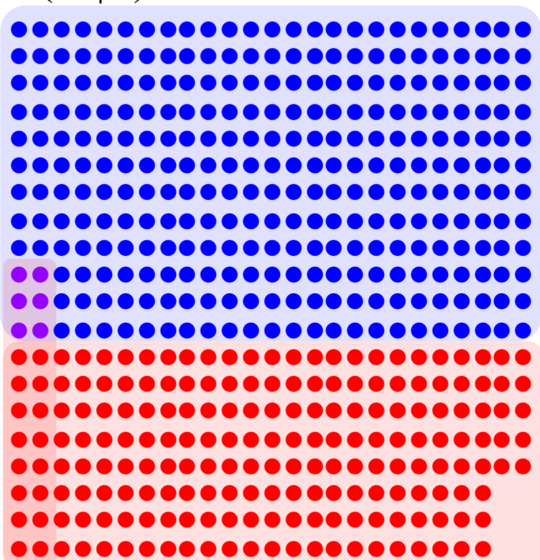
$P(B|F) \neq P(B|MF)$

That's the math definition of conditional dependence

# Conditioned on Fever

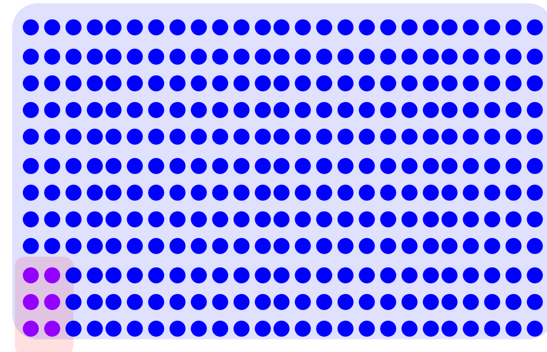
$P(B|F) = 200/494 = 0.40$
$P(M|F) = 300/494 = 0.61$

Conditioned on Fever



Test shows Malaria

Conditioned on Fever + Malaria

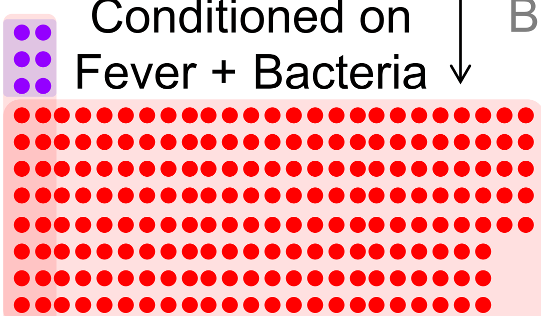


$P(B|MF) = 6/300 = 0.02$
$P(B|F) \,!\!= P(B|MF)$

That's the math definition of conditional dependence

Test shows Bacteria

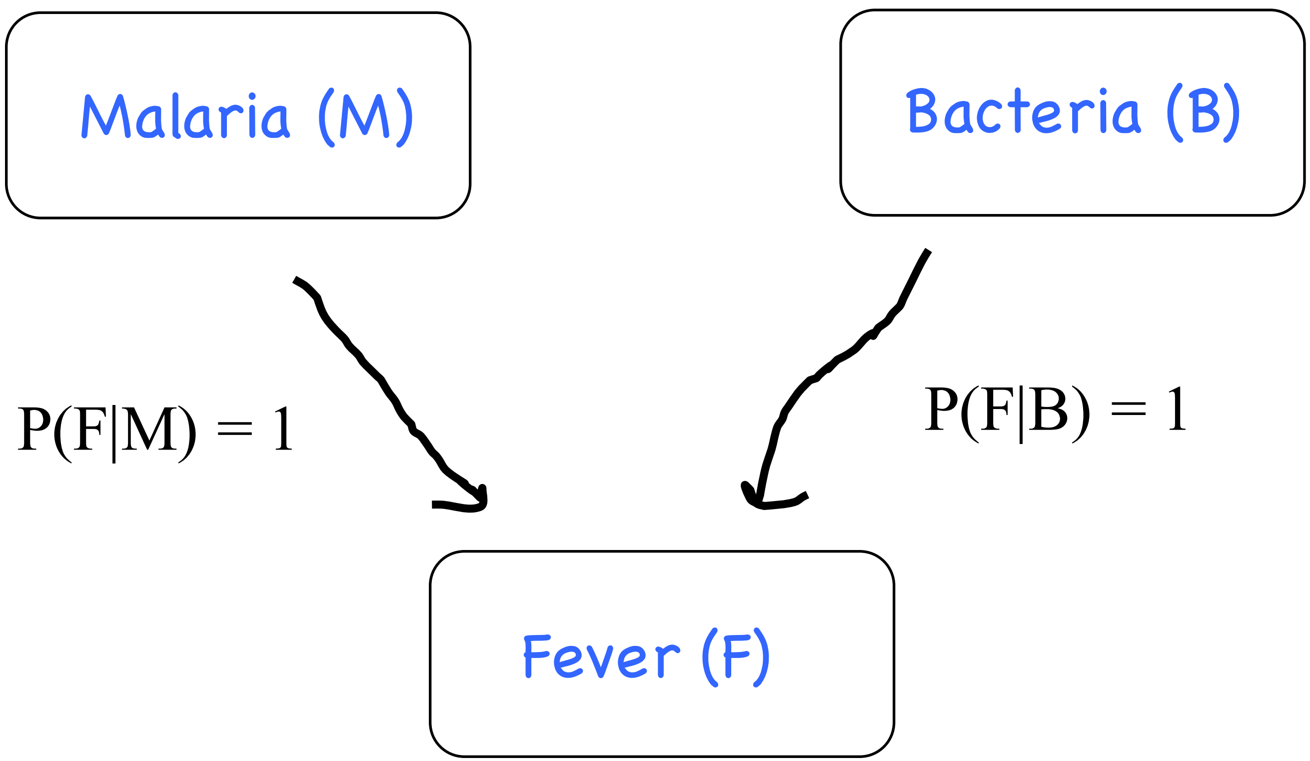
Conditioned on Fever + Bacteria



$P(M|BF) = 6/200 = 0.03$
$P(M|F) \neq P(M|BF)$

If we condition on F, the events bacteria and malaria become dependent

# Conditional Dependence



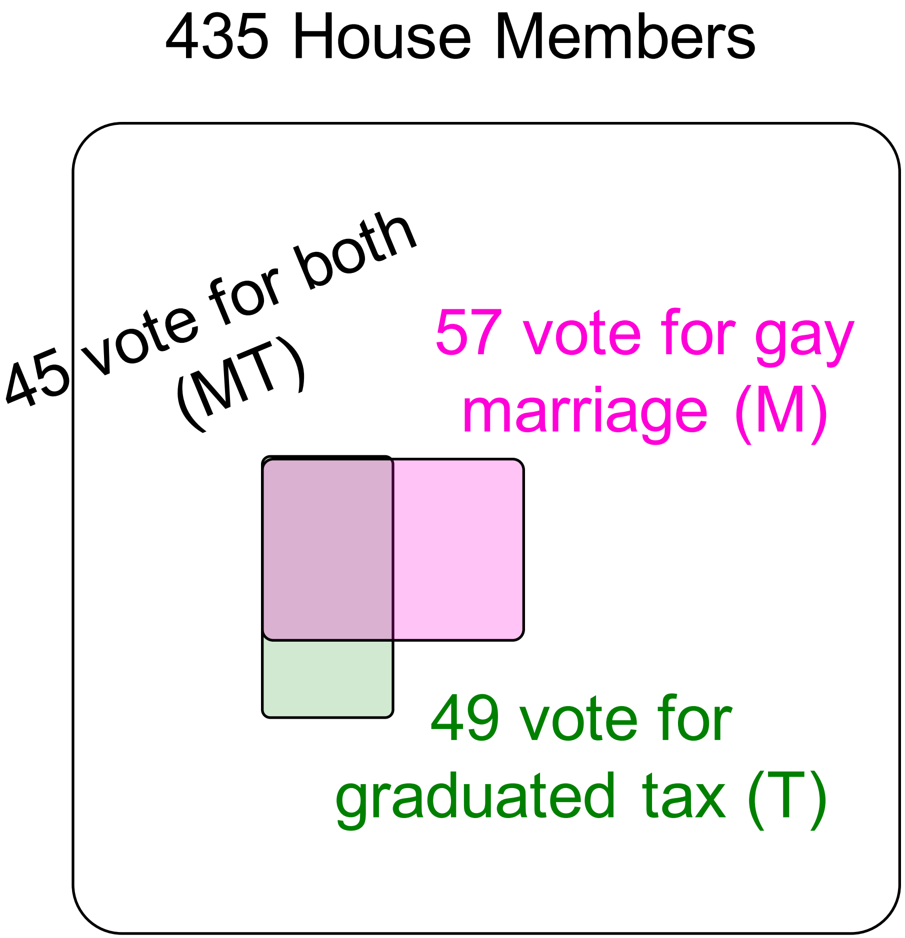

*This is a “causal” diagram. It helps explain why things are independent

# Parents With a Common Child

Say two independent parents have a common child:
When conditioned on the child they are no longer independent

# Conditional Independence

# House Voting

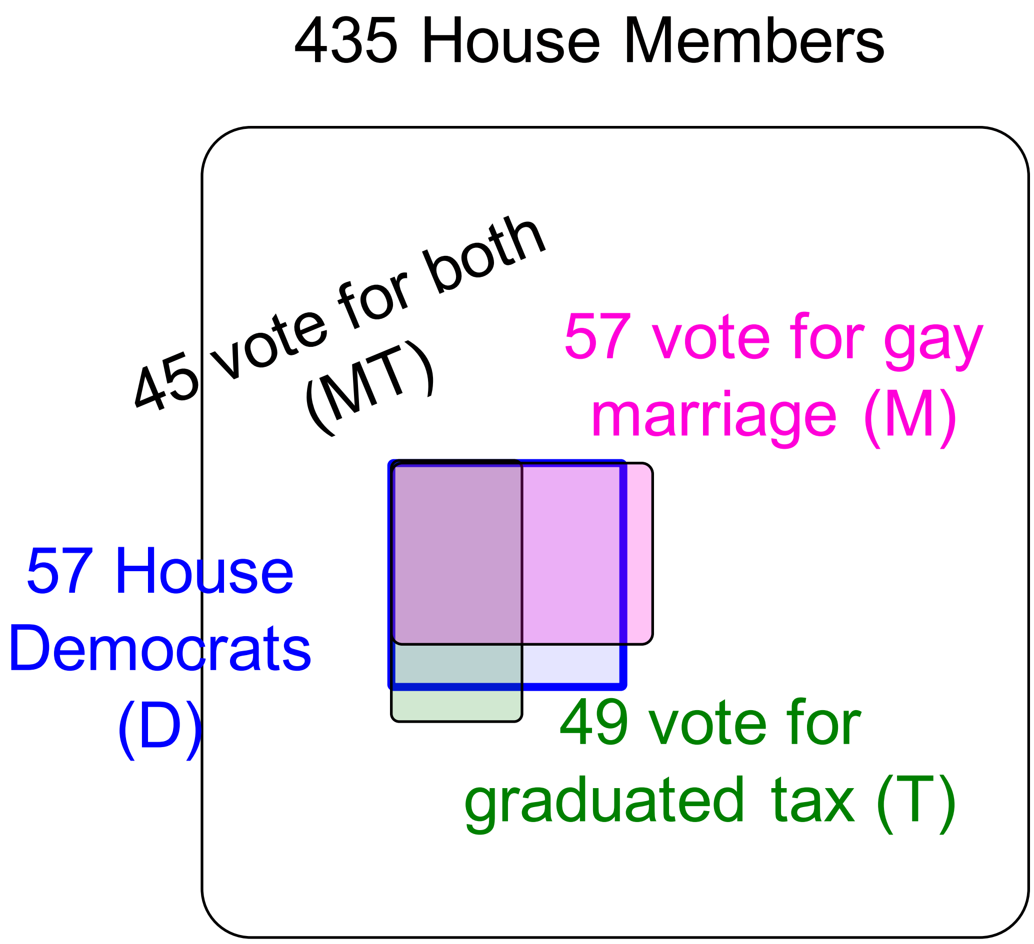



Are M and T independent?

- P(M) = 57/435 = 0.13
- P(T) = 49/435 = 0.11
- P(M)P(T) = 0.014
- P(MT) = 45/435 = 0.10

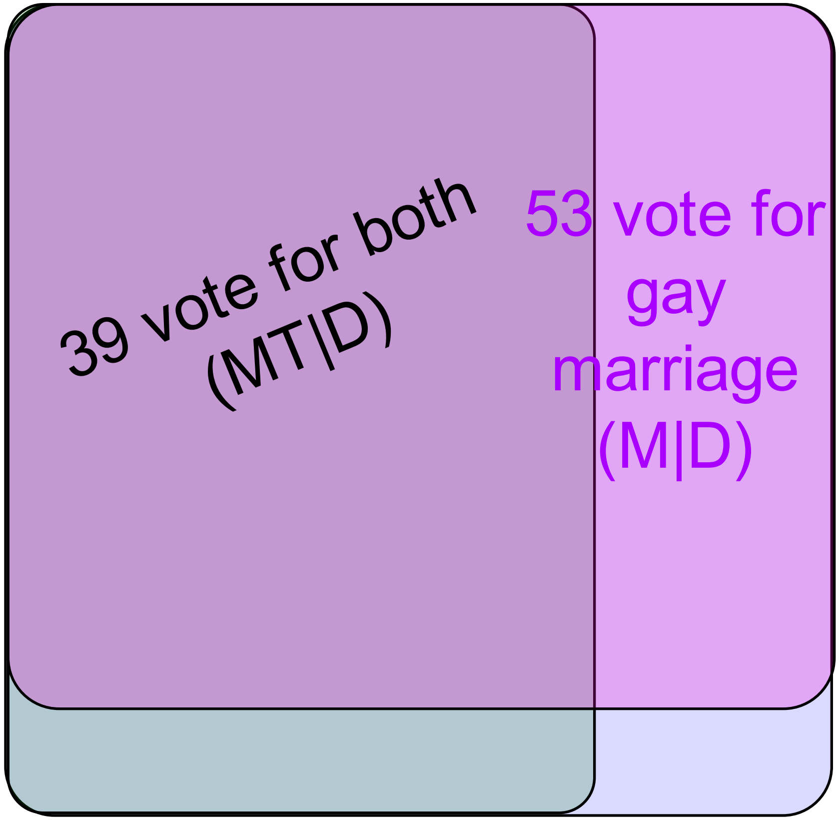
- P(MT) ≠ P(M)P(T)
- M and T are dependent

# House Voting



# Conditioned on Democrat

57 House Democrats (D)



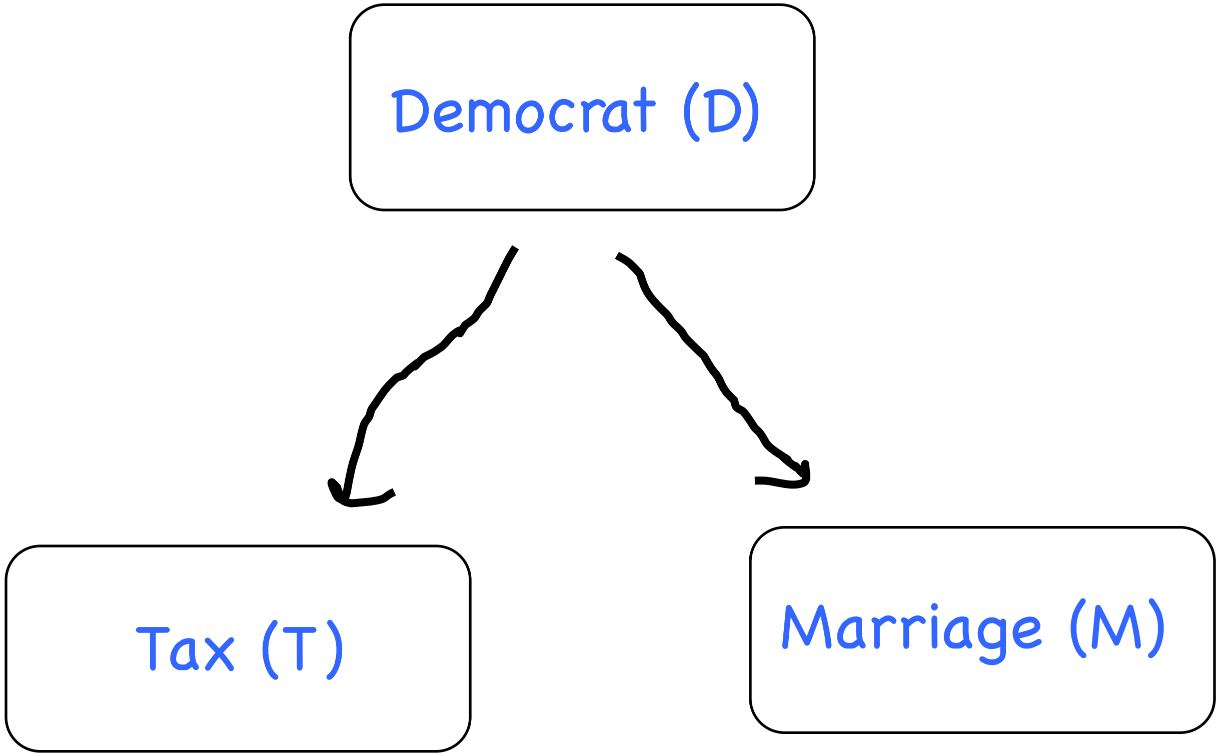

42 vote for
graduated tax (T|D)

Are M and T independent given D?

- $P(M|D) = 53/57 = 0.93$
- $P(T|D) = 42/57 = 0.74$
- $P(M|D)P(T|D) = 0.69$
- $P(MT|D) = 39/57 = 0.68$

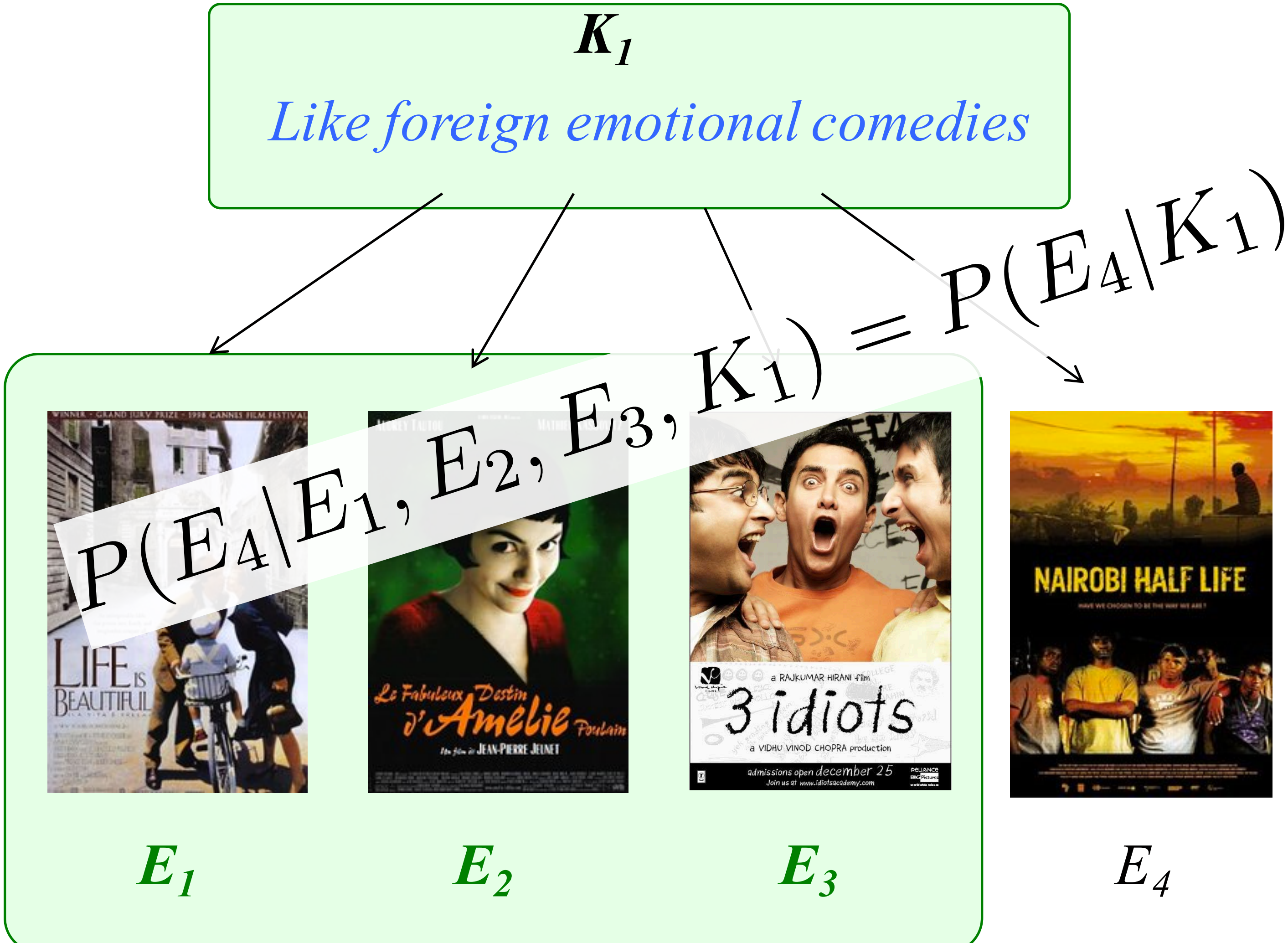
- $P(MT|D) \cong P(M|D)P(T|D)$
- M and T are conditionally independent

# Conditional Dependence




*This is a “causal” diagram. It helps explain why things are independent

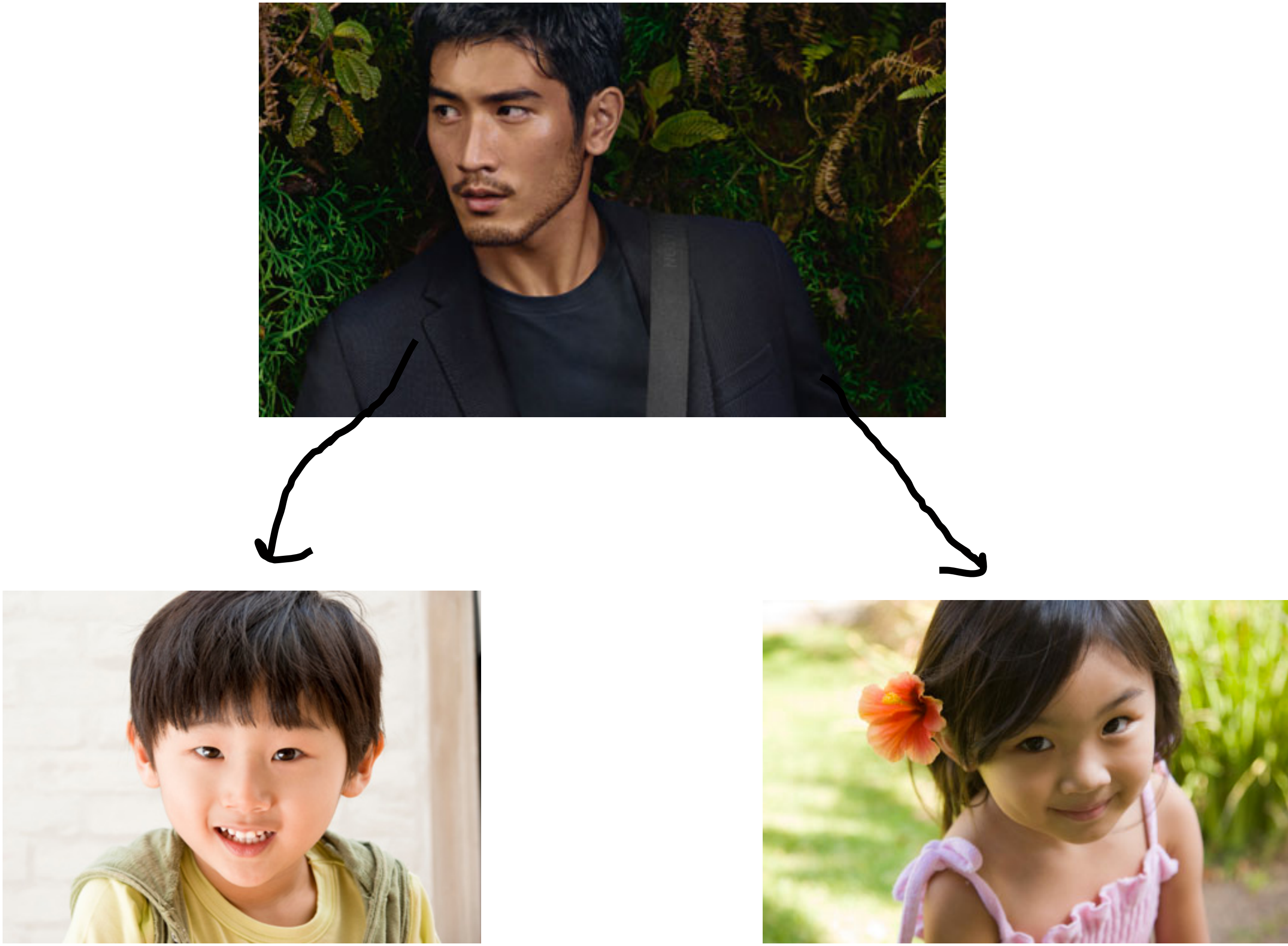
# Netflix and Learn




$$P(E_4|E_1, E_2, E_3, K_1) = P(E_4|K_1)$$

Assume $E_1$, $E_2$, $E_3$ and $E_4$ are conditionally independent given $K_1$

# Children with a Common Parent



Siblings are dependent on one another. But in the presence of a common parent,
Become independent

End Review

# Remember Learning to Code?

name

value

type

```
int a = 5;
double b = 4.2;
bit c = 1;
choice d = medium;
```

$$z \in \{\text{high}, \text{medium}, \text{low}\}$$

# Random Variable

- A **Random Variable** is a real-valued function defined on a sample space
- Example:
  - 3 fair coins are flipped.
  - Y = number of “heads” on 3 coins
  - Y is a random variable
  - $P(Y = 0) = 1/8$ (T, T, T)
  - $P(Y = 1) = 3/8$ (H, T, T), (T, H, T), (T, T, H)
  - $P(Y = 2) = 3/8$ (H, H, T), (H, T, H), (T, H, H)
  - $P(Y = 3) = 1/8$ (H, H, H)
  - $P(Y \geq 4) = 0$

# Binary Random Variable

- A binary random variable is a random variable with 2 possible outcomes (e.g., coin flip)
  - Now consider *n* coin flips, each which independently come up heads with probability *p*
  - Y = number of "heads" on *n* flips
  - P(Y = k) = $\binom{n}{k} p^k (1-p)^{n-k}$, where *k* = 0, 1, 2, …, n
  - So, $\sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = 1$
  - Proof: $\sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = (p + (1-p))^n = 1^n = 1$

# Simple Game

- Urn has 11 balls (3 blue, 3 red, 5 black)
  - 3 balls drawn. +\$1 for blue, -\$1 for red, \$0 for black
  - Y = total winnings
  - P(Y = 0) = $\left[\binom{5}{3} + \binom{3}{1}\binom{3}{1}\binom{5}{1}\right] \Big/ \binom{11}{3} = \frac{55}{165}$
  - P(Y = 1) = $\left[\binom{3}{1}\binom{5}{2} + \binom{3}{2}\binom{3}{1}\right] \Big/ \binom{11}{3} = \frac{39}{165}$ = P(Y = -1)
  - P(Y = 2) = $\binom{3}{2}\binom{5}{1} \Big/ \binom{11}{3} = \frac{15}{165}$ = P(Y = -2)
  - P(Y = 3) = $\binom{3}{3} \Big/ \binom{11}{3} = \frac{1}{165}$ = P(Y = -3)

# Probability Mass Function

- A random variable *X* is <u>discrete</u> if it has countably many values (e.g., $x_1$, $x_2$, $x_3$, ...)

- Probability Mass Function (PMF) of a discrete random variable is:

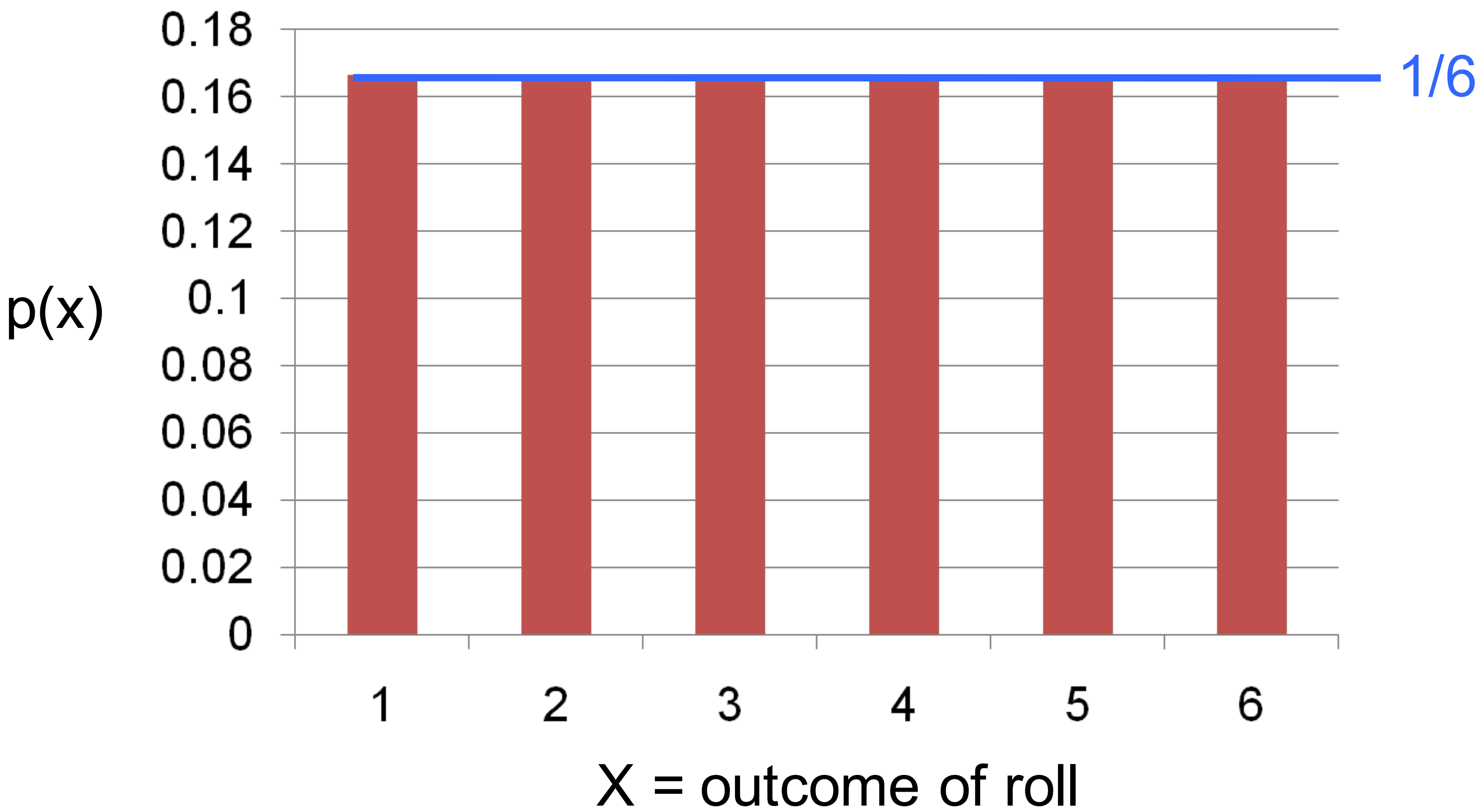
$$p(a) = P(X = a)$$

- Since $\sum_{i=1}^{\infty} p(x_i) = 1$, it follows that:

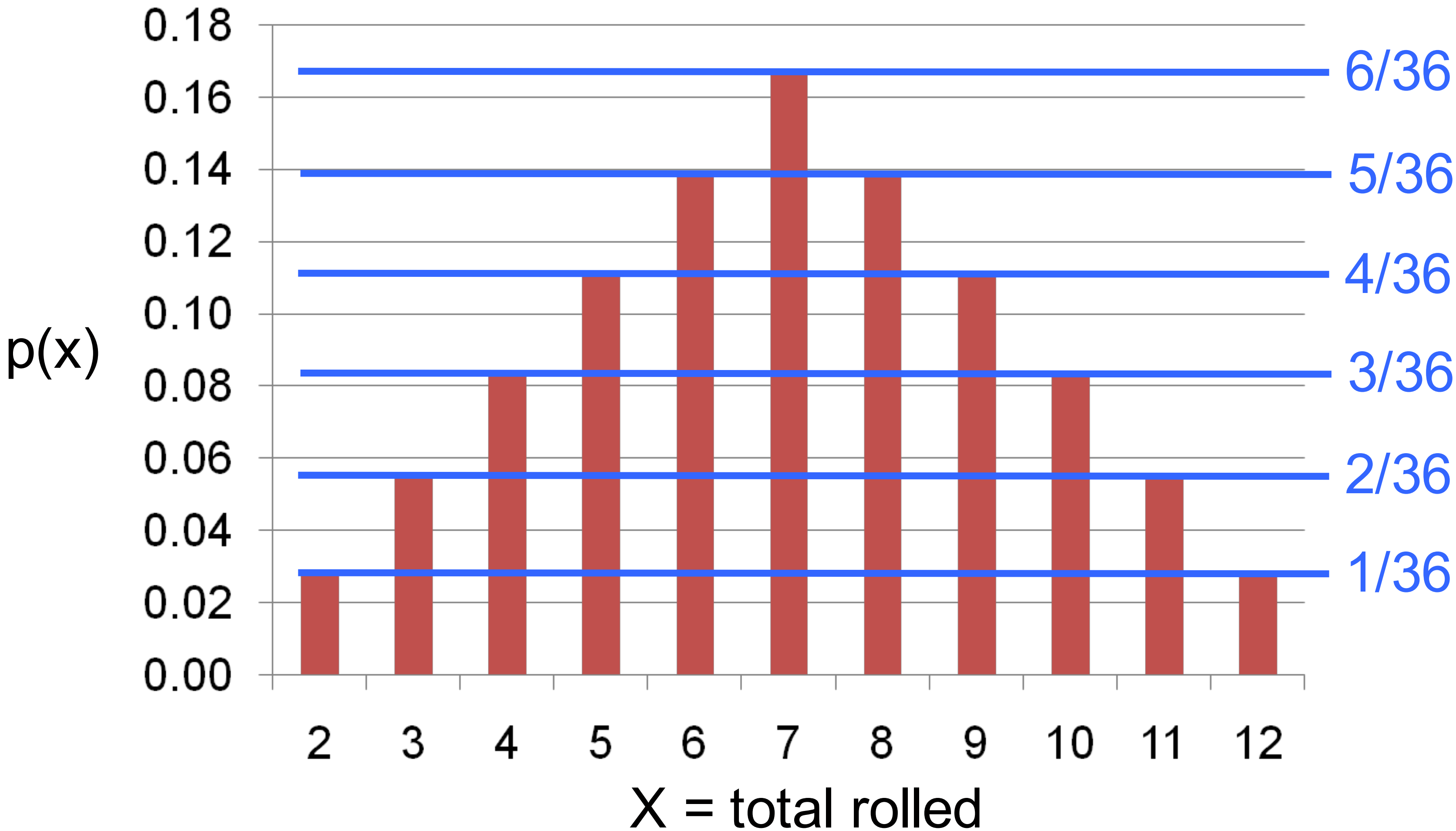
$$P(X = a) = \begin{cases} p(x_i) \geq 0 \text{ for } i = 1, 2, \ldots \\ p(x) = 0 \ \text{ otherwise} \end{cases}$$

where X can assume values $x_1$, $x_2$, $x_3$, …

# PMF For a Single 6 Sided Dice



# PMF for the sum of two dice



# Cumulative Distribution Function

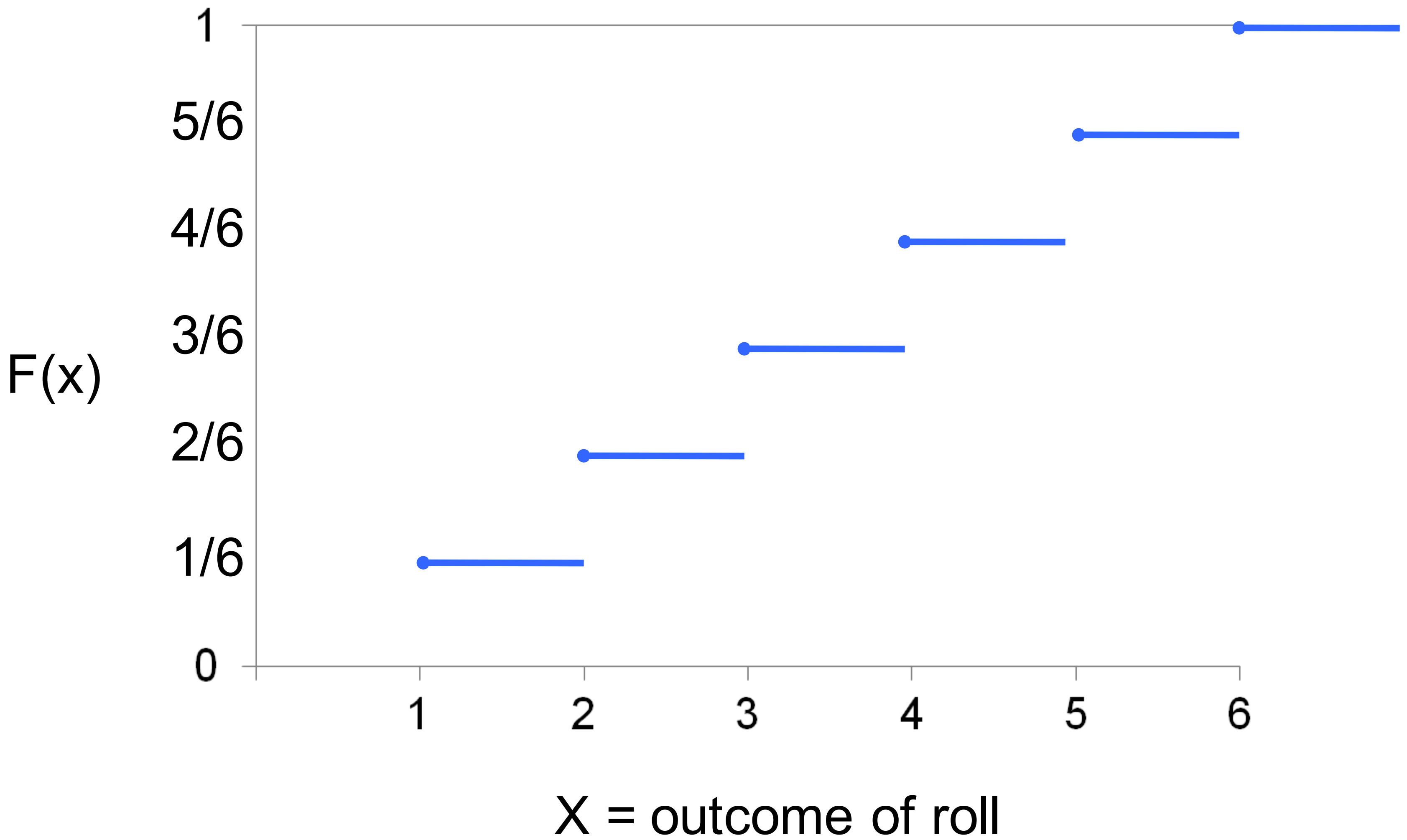
- For a random variable X, the Cumulative Distribution Function (CDF) is defined as:

$$F(a) = P(X \leq a) \quad \text{where } -\infty < a < \infty$$

- The CDF of a <u>discrete</u> random variable is:

$$F(a) = P(X \leq a) = \sum_{\text{all } x \leq a} p(x)$$

# CDF for a 6 sided dice



# Expected Value

- The Expected Values for a discrete random variable X is defined as:

$$E[X] = \sum_{x:p(x)>0} x\, p(x)$$

- Note: sum over all values of x that have p(x) > 0.

- Expected value also called: *Mean*, *Expectation*, *Weighted Average*, *Center of Mass*, *1$^{st}$ Moment*

# Expected Value

- Roll a 6-Sided Die. X is outcome of roll
  - p(1) = p(2) = p(3) = p(4) = p(5) = p(6) = 1/6
- E[X] = $1\left(\frac{1}{6}\right)+2\left(\frac{1}{6}\right)+3\left(\frac{1}{6}\right)+4\left(\frac{1}{6}\right)+5\left(\frac{1}{6}\right)+6\left(\frac{1}{6}\right)=\frac{7}{2}$
- Y is random variable
  - P(Y = 1) = 1/3, P(Y = 2) = 1/6, P(Y = 3) = 1/2
- E[Y] = 1 (1/3) + 2 (1/6) + 3 (1/2) = 13/6

# Indicator Variable

- A variable $I$ is called an indicator variable for event *A* if

$$I = \begin{cases} 1 & \text{if } A \text{ occurs} \\ 0 & \text{if } A^c \text{ occurs} \end{cases}$$

- What is E[$I$]?
  - p($I$ =1) = P(A), p($I$ = 0) = 1 – P(A)
  - E[$I$] = 1 P(A) + 0 (1 – P(A)) = P(A)

*We'll use this property frequently!*

# Lying with Statistics

"There are three kinds of lies:
lies, damned lies, and statistics"

– *Mark Twain*

- School has 3 classes with 5, 10 and 150 students
- Randomly choose a class with equal probability
- X = size of chosen class
- What is E[X]?
  - E[X] = 5 (1/3) + 10 (1/3) + 150 (1/3)
    = 165/3 = 55

# Lying with Statistics

"There are three kinds of lies:
lies, damned lies, and statistics"

– *Mark Twain*

- School has 3 classes with 5, 10 and 150 students
- Randomly choose a student with equal probability
- Y = size of class that student is in
- What is E[Y]?
  - $E[Y] = 5\ (5/165) + 10\ (10/165) + 150\ (150/165)$
    $= 22635/165 \approx 137$
- Note: E[Y] is students' perception of class size
  - But E[X] is what is usually reported by schools!

# Expectation of a Function

- Let Y = g(X), where g is real-valued function

$$E[g(X)] = E[Y] = \sum_j y_j p(y_j)$$

$$= \sum_j y_j \sum_{i:g(x_i)=y_j} p(x_i)$$

$$= \sum_j \sum_{i:g(x_i)=y_j} y_j \; p(x_i)$$

$$= \sum_j \sum_{i:g(x_i)=y_j} g(x_i) p(x_i)$$

$$= \sum_i g(x_i) p(x_i)$$

# Properties of Expectation

- Linearity:

$$E[aX + b] = aE[X] + b$$

  - Consider X = 6-sided die roll, Y = 2X – 1.
  - E[X] = 3.5 E[Y] = 6

- *N*-th Moment of X:

$$E[X^n] = \sum_{x : p(x) > 0} x^n p(x)$$

  - We'll see the 2nd moment soon...