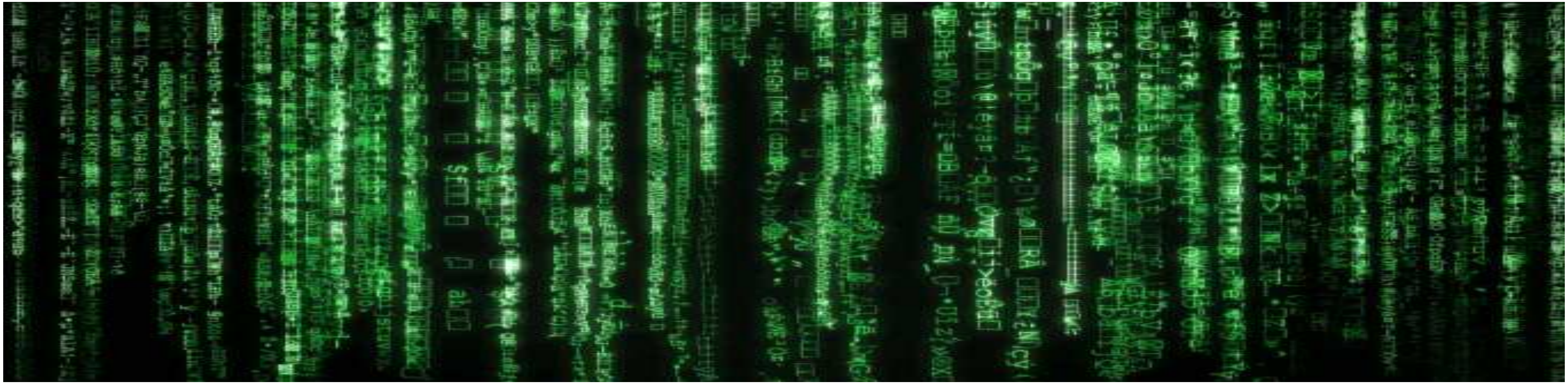


INF 110 Discovering Informatics

Probability

NAU
NORTHERN
ARIZONA
UNIVERSITY

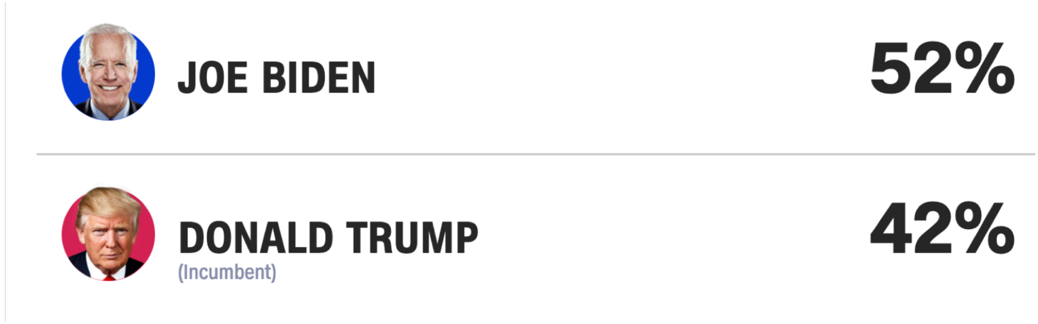
Introduction

The principles of probability are everywhere.

People use the term probability many times each day.

For example, physician says that a patient has a 50-50 chance of surviving a certain operation. Another physician may say that she is 95% certain that a patient has a particular disease

CNN Polling Averages from 2020 US Presidential Election:

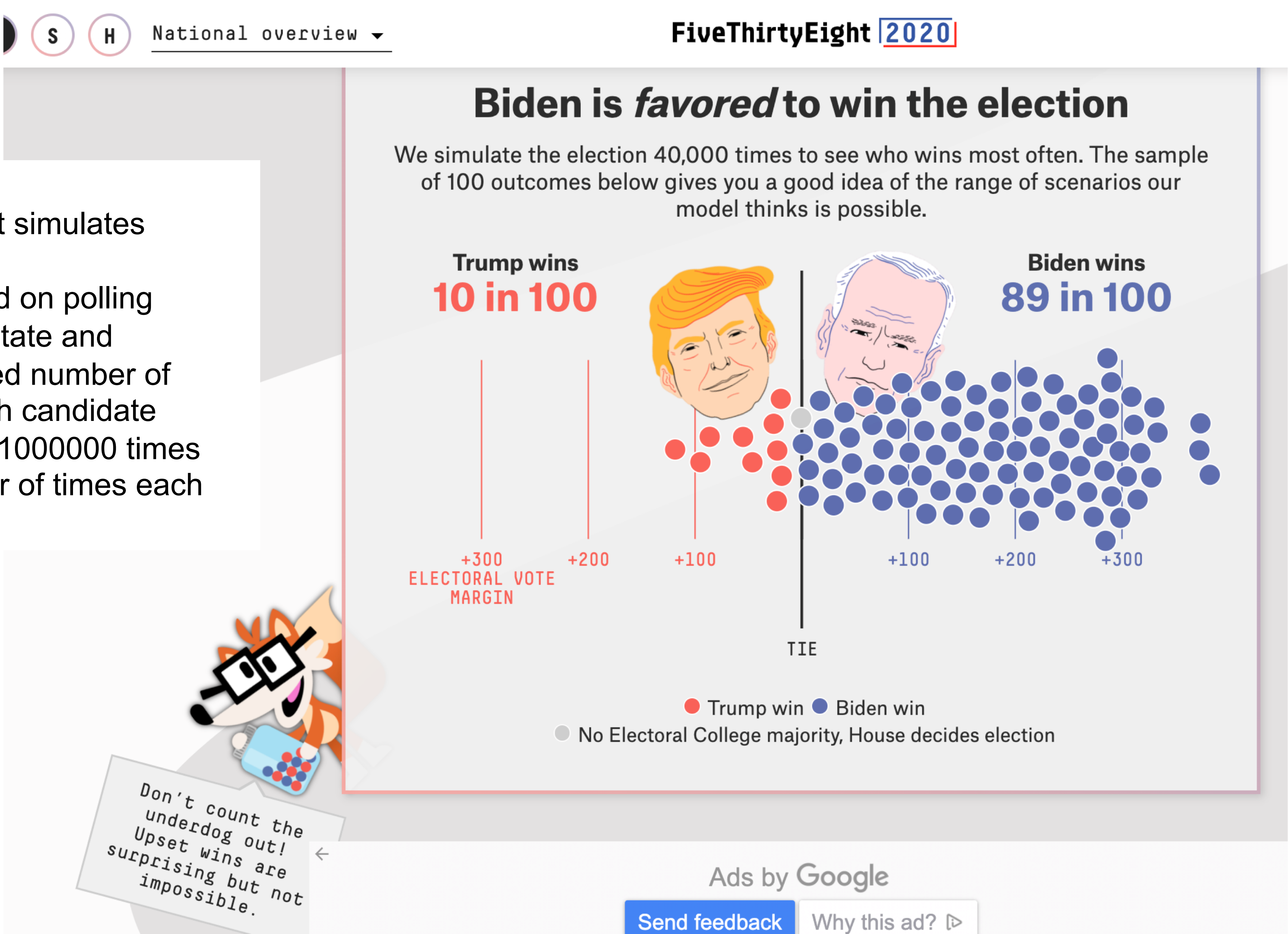


Election Results:

- US voter preference polls have underestimated President Donald Trump’s eventual proportion of total votes
- Polls aren’t the same as election results – although they are often used to predict elections
- This is where probabilistic error comes in
  - Sampling methods (phone, internet)
  - Likely voter models
  - Polls that were taken closer to election day are more accurate
  - campaigns release their own polling (sus)

Election Forecasting:
538 creates a model that simulates election results

- The models are based on polling averages from each state and estimates the expected number of electoral votes to each candidate
- They run the model ~1000000 times and report the number of times each candidate wins



Election Forecasting:

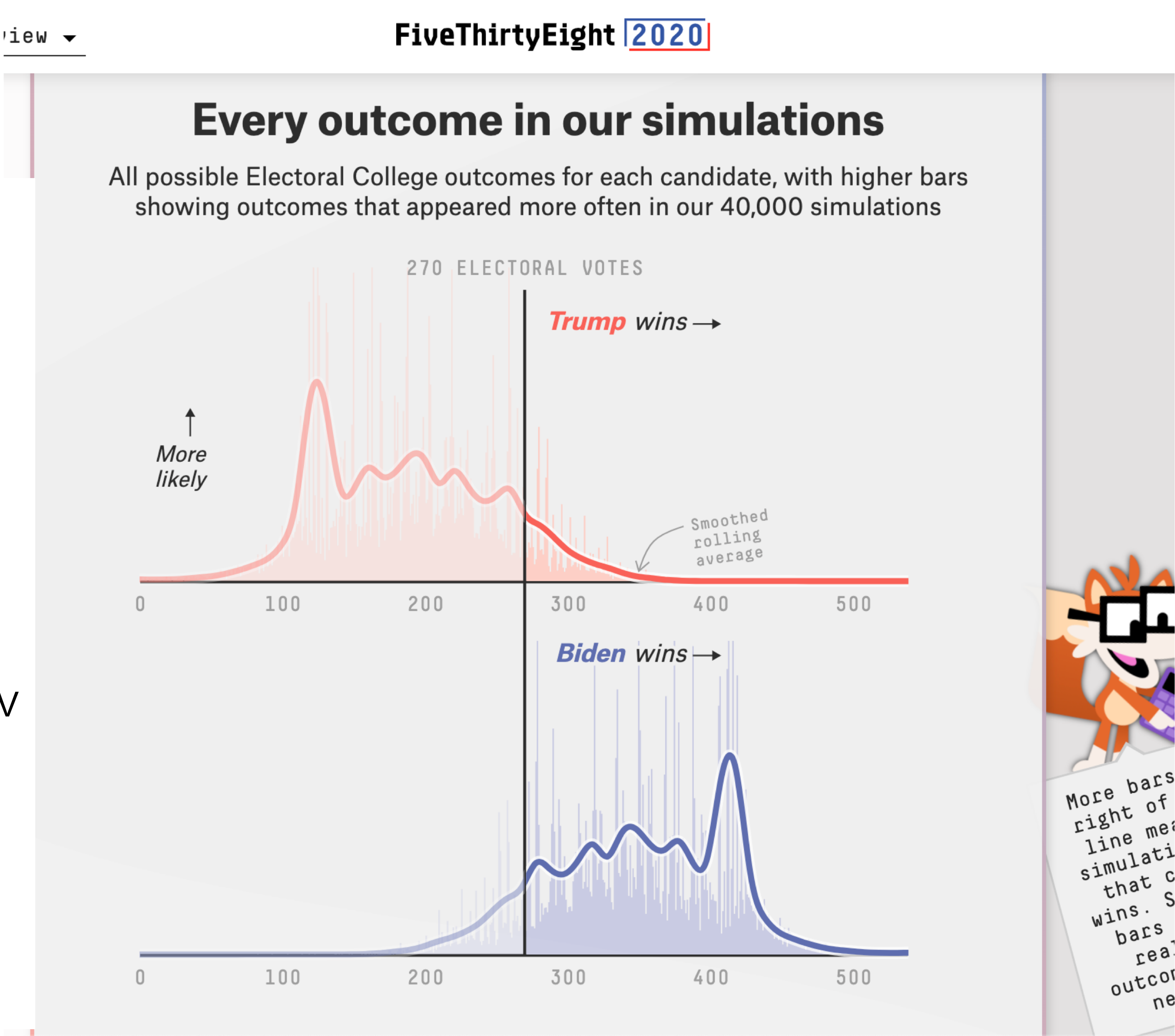
538 creates a model that simulates election results

- The models are based on polling averages from each state and estimates the expected number of electoral votes to each candidate
- They run the model ~1000000 times and report the number of times each candidate wins

The EV tally in each outcome is recorded

- Many outcomes include Biden >400 EV
- But also a lot around Biden ~300 EV

In reality, polling errors meant that because of some very close states, there were more plausible scenarios where Trump could win, and the election was closer (306-232)



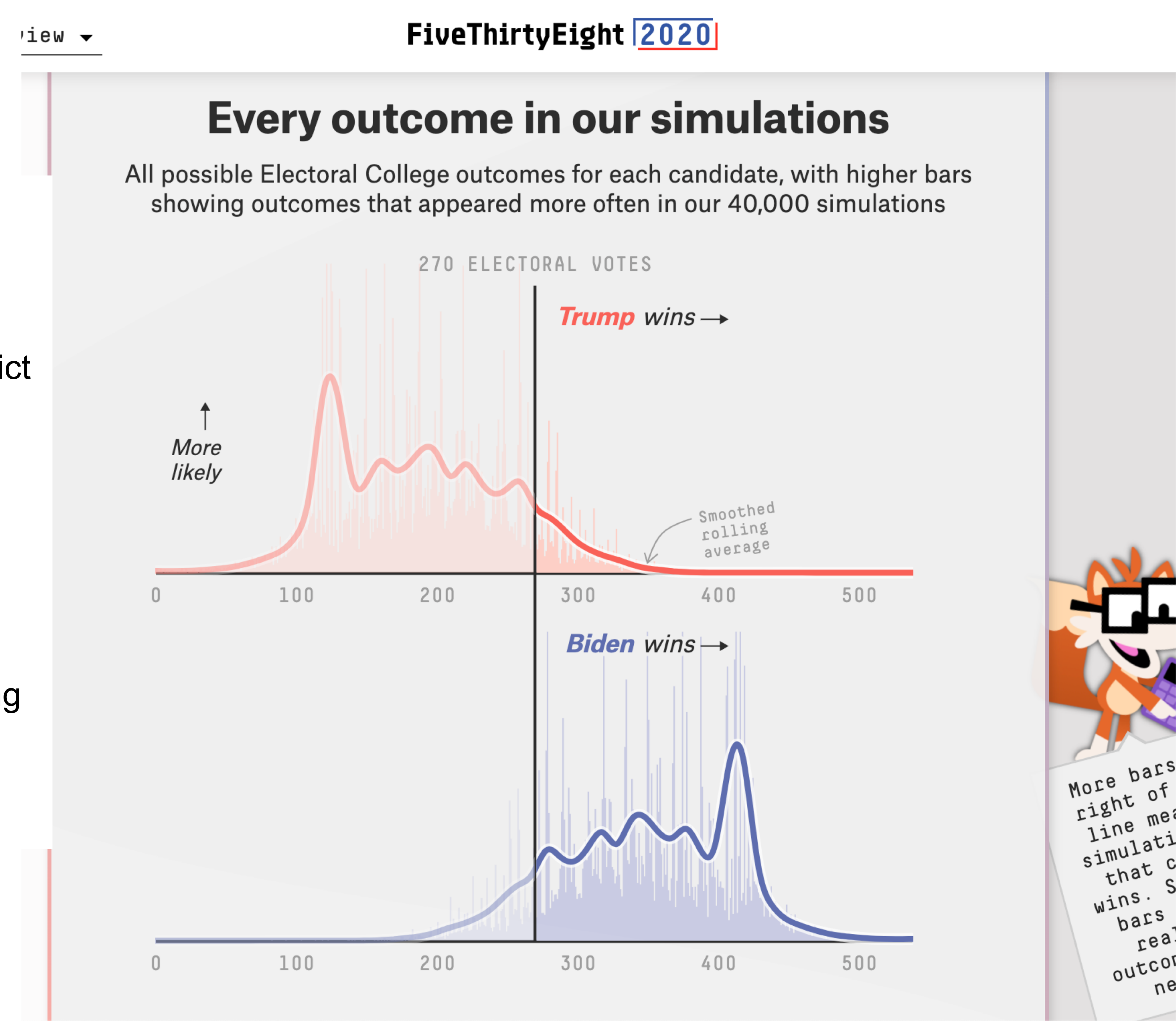
Election Forecasting:
IMO, election forecasts are a product being sold to media and advertisers

They are useful because they help predict an outcome

They are generally correct – serious statisticians and programmers work on them

They are predictions based on probabilities and suffer from the sampling biases underlying all polling techniques

They are not fortunetellers



Definition of Probability

If an event can occur in **N** mutually exclusive and equally likely ways, and if **m** of these possess a trait, **E**, the probability of the occurrence of E is read as

$$P(E) = m/N$$

Example: Coin Flips

- We know a fair coin has a 50/50 probability of heads or tails
- But in terms of our definition of probability:
  - N = 2 ways a fair coin can appear
  - m = 1 configuration where heads appears

P(Heads) = ½ = 0.5 = 50%

Example: More Dice (d20)

- N = 20 ways a fair d20 can appear
- m = 1 configuration where 20 appears

P(Twenty) = 1/20 = 0.05 = 5%

Random Selection

First step in calculating probability computationally.

We can choose a random value (like a dice roll) with numpy:

```
np.random.choice(set, number=1)
```

Random Selection

First step in calculating probability computationally.

We can choose a random value (like a dice roll) with numpy:

```
np.random.choice(set, number=1)
```

set of all possible choices

number of random selections from the set

Live Code Rolling Dice

Tasks: "Roll" a d20 in Python; Then roll a d20 100 times!

Learning Outcomes

- Using the choice method
- Understanding random sets

Definitions

- ***Sample space***: collection of unique, non-overlapping possible outcomes of a random circumstance.
  - i.e., “heads” and “tails”
- ***Simple event:*** one outcome in the sample space; a possible outcome of a random circumstance.
  - must be defined within the sample space
  - i.e., when rolling a 6-sided die, “20” is not a simple event
- ***Event:*** not necessarily a single event
  - i.e., what is the probability that we roll a 2 and a 4 in 3 rolls?

Live Code Rolling Weighted Dice

Tasks: According to the World Health Organization (WHO) there are 105 male births for every 100 female births. The events in the sample space aren't equally likely. But we can treat them like they are if we change the sample space from 2 to 205.

P(M) = 0.512;   P(F) = 0.488

Learning Outcomes

- Using the choice method
- Understanding random sets

Definitions

- ***Complement*** ==> sometimes, we want to know the probability that an event will not happen; an event opposite to the event of interest is called a complementary event.
- If A is an event, its complement is The probability of the complement is AC or !A
- Example: The complement of heads event is tails

- $P(A) + P(AC) = 1$

Views of Probability

Subjective

- It is an estimate that reflects a person’s opinion, or best guess about whether an outcome will occur.
- This is a common view in medicine: probability forms the basis of a physician’s opinion about whether a patient has a specific disease.
  - i.e., “according to the test you have an 80% chance of…”
- Estimate may change based on diagnostic procedures.

Views of Probability

Objective

- More formally verifiable than subjective probability.
- It is well known that the probability of flipping a fair coin and getting tails is 0.50.
- If a coin is flipped 10 times, is there a guarantee, that exactly 5 tails will be observed?
- If the coin is flipped 100 times? With 1000 flips?
- As the number of flips becomes larger, the proportion of coin flips that result in tails approaches 0.50

Live Code Flipping Coins

Tasks: Flip a coin 10x, 100x, 1000x and count the # of heads.
What happens as the number of flips grows larger?

Hint: To count the number of heads, use:

```
np.count_nonzero(rolls == "T")
```

Learning Outcomes

- Using the choice method
- Understanding random sets

Views of Probability

Objective

- Relative frequency of occurrence
- It’s important to conduct a sufficient number of trails
- Assuming that an experiment can be repeated many times and assuming that there are one or more outcomes that can result from each repetition.
- Then, ***the probability of a given outcome is the number of times that outcome occurs divided by the total number of repetitions***.

Example Using Data

- P(Male and AB)?
- P(AB)?

Blood Group	Males	Females	Total
O A B AB	20 17 8 5	20 18 7 5	40 35 15 10
Total	50	50	100

Marginal Probabilities

Named because they appear on the “margins” of a probability table. It is the probability of a ***single outcome***

Examples: P(Male), P(AB)

Conditional Probabilities

It is the probability of an event on condition that certain criteria is satisfied

Example: If a subject was selected randomly and found to be female what is the probability that she has a blood group O or P(O\F)

P(O\F) = 20/50 = 0.40

the back slash means "probability of O given F"

Joint Probability

It is the probability of occurrence of two or more events together

While conditional probability asks given that the subject is male, what is the probability he has type AB, joint probability asks if we pick a subject at random from the sample space, what is the probability that the subject is male and has AB?

Example: Probability of being male & belong to blood group AB

$$P(M \text{ and } AB) = P(M \cap AB) = 5/100 = 0.05$$

$\cap$ = intersection

Joint Probability

$\cap$ = intersection

Comes from ***set theory notation***

Portion shared by both sets

Think of the shared area of a Venn diagram

(pronounced “AND”)

Probability Basics

- Probabilities ranges between 0 and 1
- If an outcome cannot occur, its probability is 0
- If an outcome is sure, it has a probability of 1
- The sum of probabilities of mutually exclusive outcomes is equal to 1

$$P(M) + P(F) = 1$$

Multiplication Rule

If two events are independent, then the probability of them ***both*** occurring is the probability of the first event ***TIMES*** the probability of the second event.

P(A and B) = P(A) P(B)

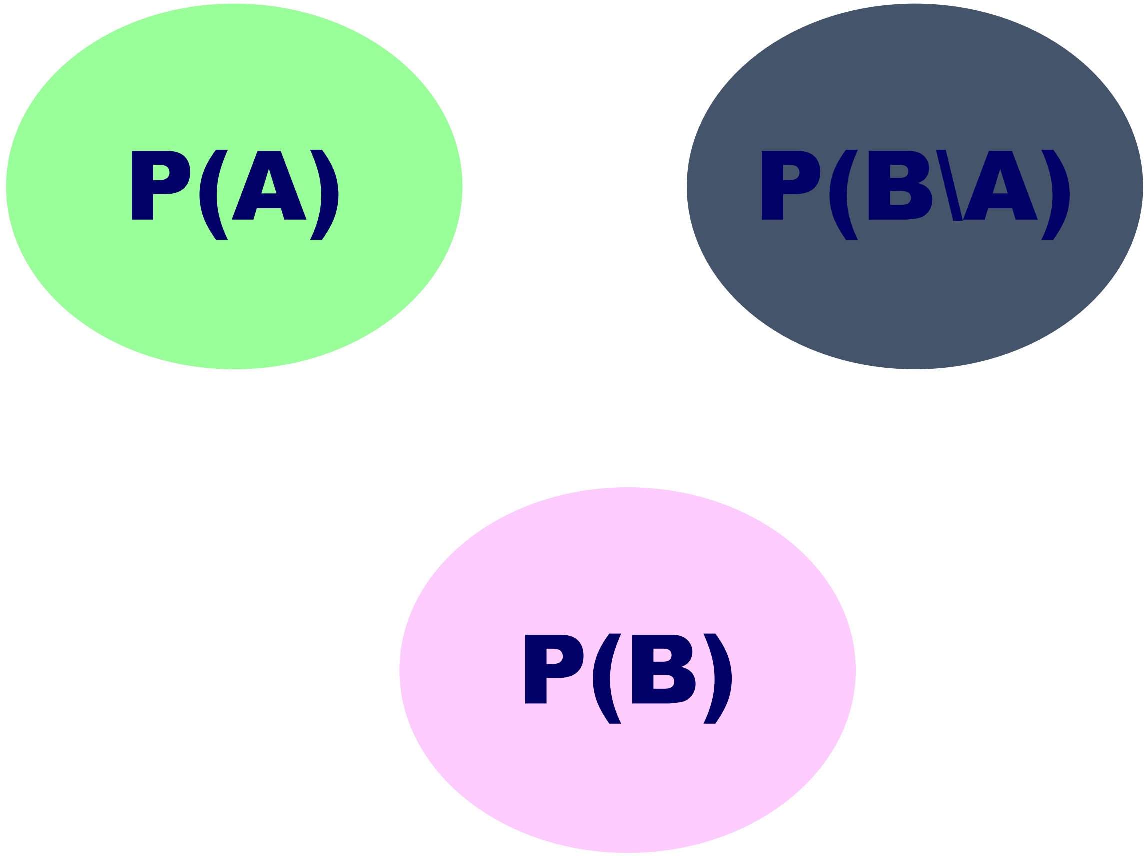
i.e., what’s the probability of flipping a coin twice and getting heads both times?

How do you know if two events are independent?

A and B are independent if P(B\A) = P(B)

If we learn about the occurrence of A does it give any information about the occurrence of B?

If not, then the events are independent.



Example

Consider the joint probability of being male and having blood type O

P(O) = 40/100 = 0.40

P(O\M) = 20/50 = 0.40

The two events are independent! So we can calculate the probability of being male and having blood type O:

$P(O \cap M) = P(O) \times P(M) = (40/100) \times (50/100) = 0.20$

Multiplication Rule (When Events are **NOT** Dependent)

Instead of multiplying straight through, we account for dependency

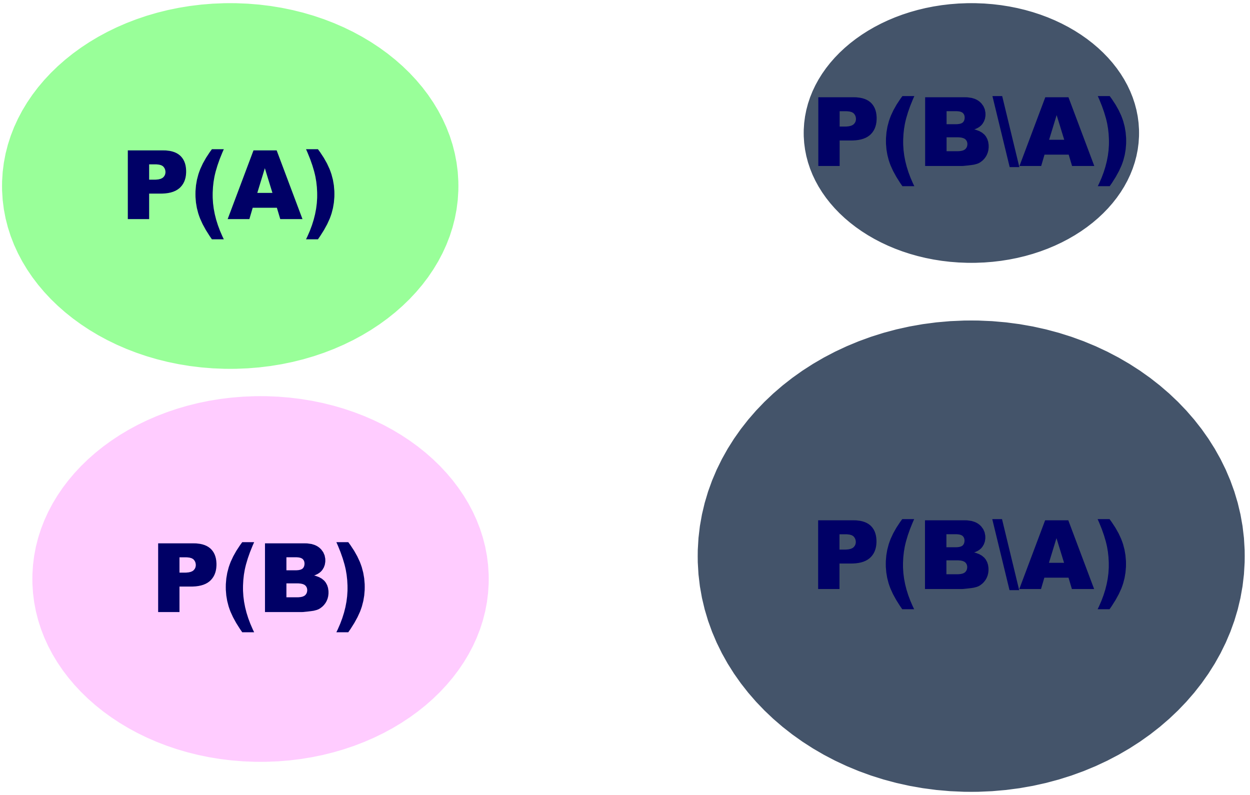
$\boxed{P(A \text{ and } B)} = \boxed{P(A)}\ \boxed{P(B \backslash A)}$

We're using the ***marginal probability*** of A times the ***conditional probability*** of B given A to find the ***joint probability*** of (A and B)

How do you know if two events are dependent?

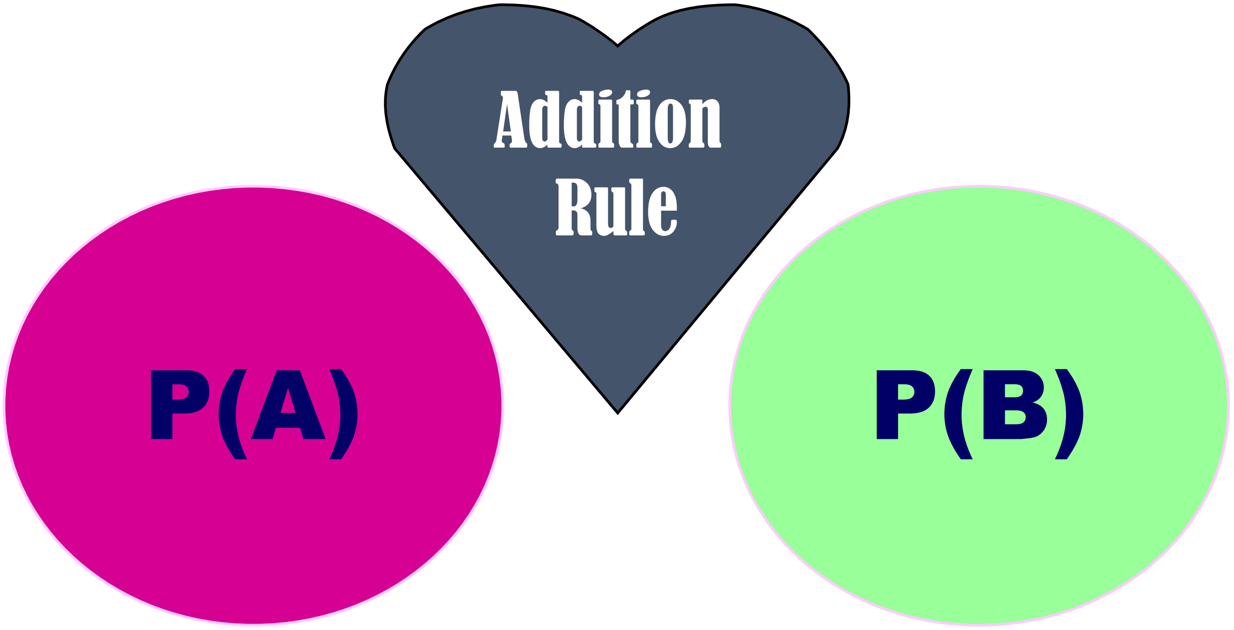
This is essentially the complement of the previous rule.

A and B are dependent if P(B) != P(B\A)



Rules of Probability: Addition

- A and B are mutually exclusive
- The occurrence of one event precludes the occurrence of the other
- The probability of A OR B is the sum of the two probabilities.



P(A OR B) = P(A U B) = P(A) + P(B)

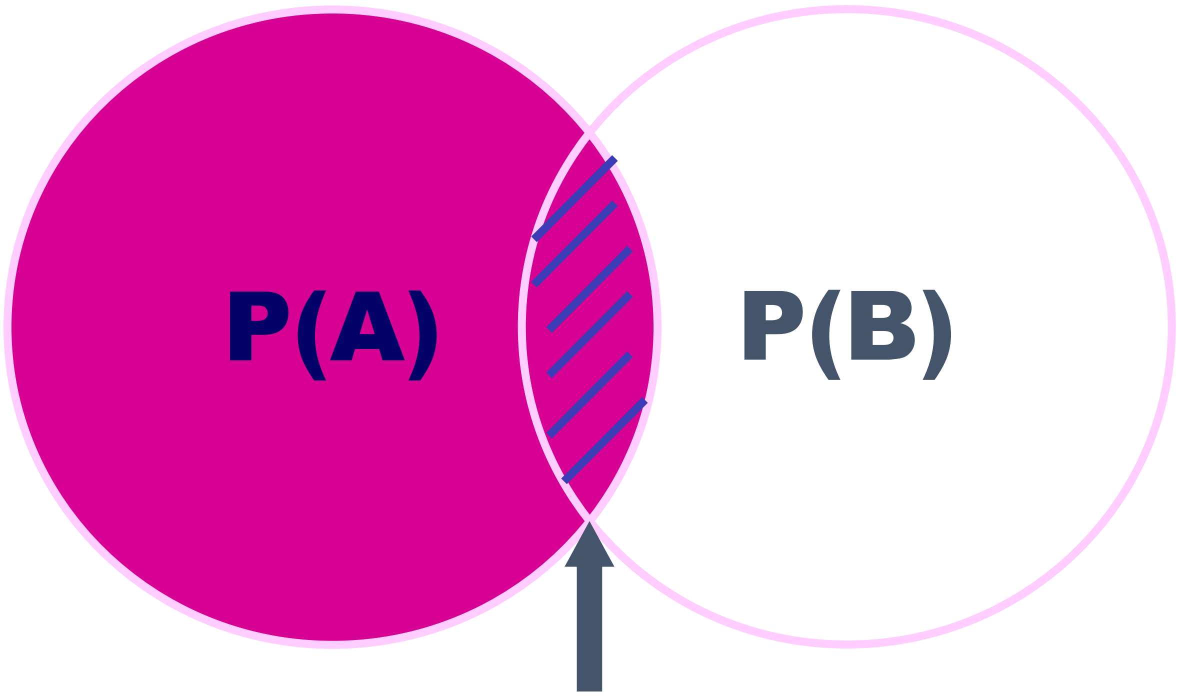
Example

The probability of being either blood type O or blood type A:

$$P(O \cup A) = P(O) + P(A)$$
$$= (40/100)+(35/100)$$
$$= 0.75$$

A and B are non mutually exclusive
(Can occur together)
Example: Male and smoker

Modified Addition Rule



$P(A \cap B)$ Subtract this intersection so we don't count it twice

$P(A \text{ OR } B) = P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Example

Two events (male gender and blood type O) are not mutually exclusive.

$$\begin{aligned} P(\text{M OR O}) &= P(M)+P(O) - P(M \cap O) \\ &= 0.50 + 0.40 - 0.20 \\ &= 0.70 \end{aligned}$$

end