







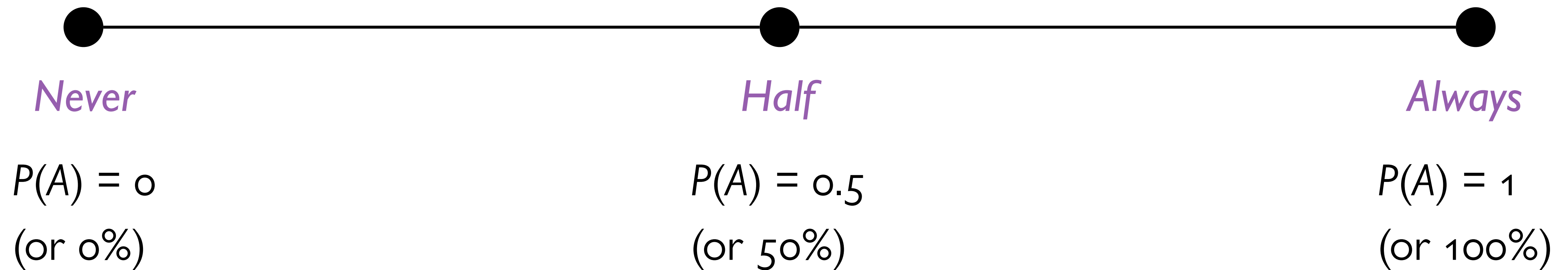
*Roman coin featuring  
Pompey the Great and a  
ship*

If you flip a fair coin 100 times, *what is the chance* of getting between 40 and 60 heads?

***The intuition of probability:*** *If we repeat an event (flipping a coin) many times (say, 1 million) ... what fraction of the times would we see the outcome we care about (heads)?*

# Probability

$P(A)$ : Chance that an event  $A$  will occur



$$P(\text{heads}) = 1/2$$

# How to calculate an event's probability?

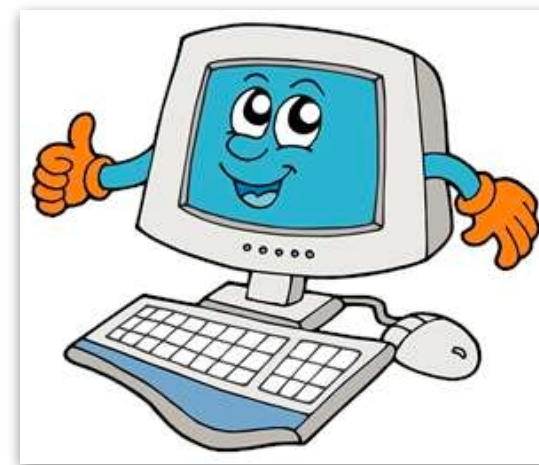
## Computers (simulation)

```
N = 1000000 #Roll the dice 1 million times
option_a = np.random.choice(dice, N) + np.random.c
option_b = 2 * np.random.choice(dice, N)

print("Option A Mean: ", np.mean(option_a))
print("Option B Mean: ", np.mean(option_b))

Option A Mean: 7.003198
Option B Mean: 6.99884

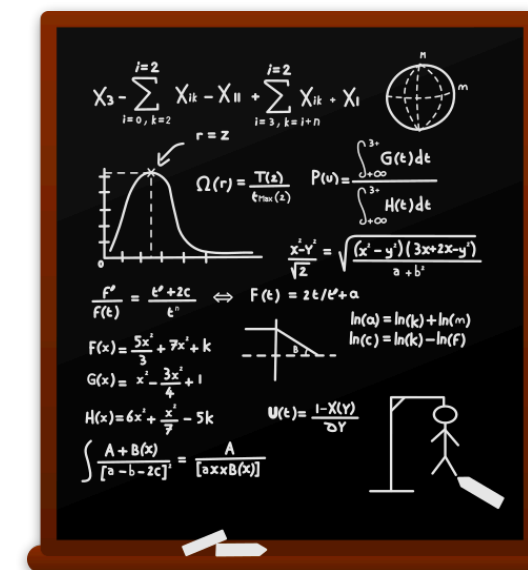
samples = Table().with_columns("Option A", option_
samples.hist("Option A", bins=np.arange(0,14))
samples.hist("Option B", bins=np.arange(0,14))
```



Rooted in *algorithms*

- ✗ Approximate solutions
- ✓ Often convincing
- ✓ Non-trivial problems potentially captured cleanly with code

## Math (analytical)



Rooted in *rules (axioms)*

- ✓ Exact solutions
- ✓ Straightforward for simple problems
- ✗ Non-trivial problems potentially difficult to analyze/verify

# How to calculate an event's probability?

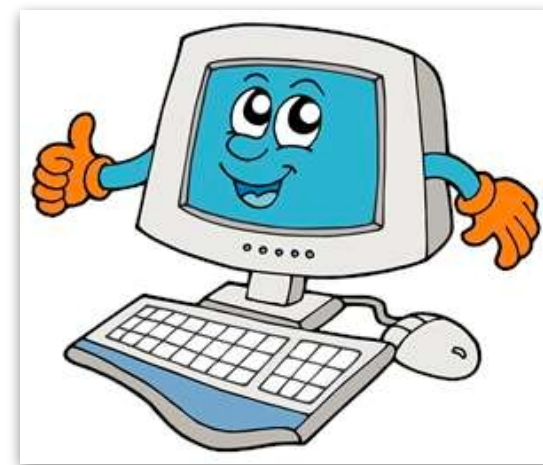
## Computers (simulation)

```
N = 1000000 #Roll the dice 1 million times
option_a = np.random.choice(dice, N) + np.random.c
option_b = 2 * np.random.choice(dice, N)

print("Option A Mean: ", np.mean(option_a))
print("Option B Mean: ", np.mean(option_b))

Option A Mean: 7.003198
Option B Mean: 6.99884

samples = Table().with_columns("Option A", option_
samples.hist("Option A", bins=np.arange(0,14))
samples.hist("Option B", bins=np.arange(0,14))
```



Rooted in *algorithms*

- ✗ Approximate solutions
- ✓ Often convincing
- ✓ Non-trivial problems potentially captured cleanly with code

Repeat many times:

Simulate one trial

Record the outcome

Analyze outcomes for all trials

# How to calculate an event's probability?

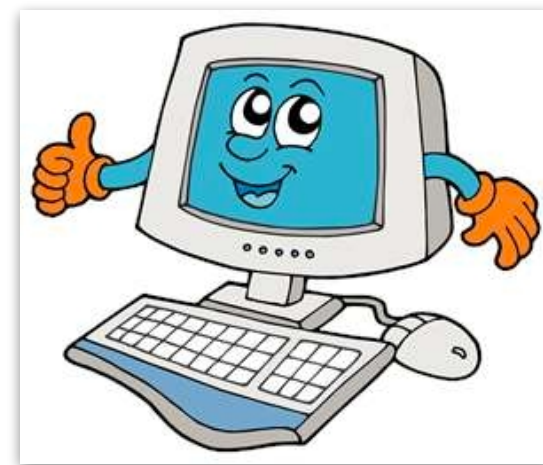
## Computers (simulation)

```
N = 1000000 #Roll the dice 1 million times
option_a = np.random.choice(dice, N) + np.random.c
option_b = 2 * np.random.choice(dice, N)

print("Option A Mean: ", np.mean(option_a))
print("Option B Mean: ", np.mean(option_b))

Option A Mean: 7.003198
Option B Mean: 6.99884

samples = Table().with_columns("Option A", option_
samples.hist("Option A", bins=np.arange(0,14))
samples.hist("Option B", bins=np.arange(0,14))
```



Rooted in *algorithms*

- ✗ Approximate solutions
- ✓ Often convincing
- ✓ Non-trivial problems potentially captured cleanly with code

Repeat many times:

*Flip 100 coins*

Record the outcome

Analyze outcomes for all trials

# How to calculate an event's probability?

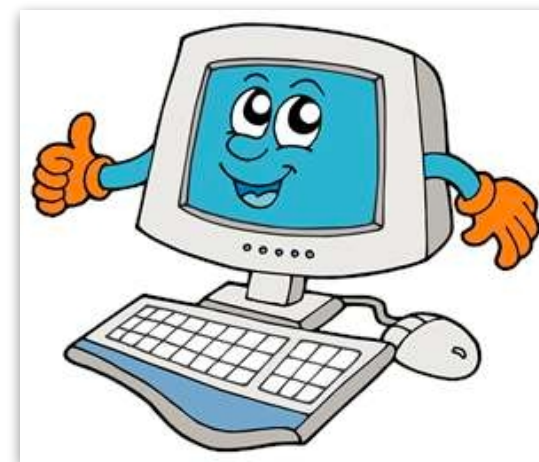
## Computers (simulation)

```
N = 1000000 #Roll the dice 1 million times
option_a = np.random.choice(dice, N) + np.random.c
option_b = 2 * np.random.choice(dice, N)

print("Option A Mean: ", np.mean(option_a))
print("Option B Mean: ", np.mean(option_b))

Option A Mean: 7.003198
Option B Mean: 6.99884

samples = Table().with_columns("Option A", option_
samples.hist("Option A", bins=np.arange(0,14))
samples.hist("Option B", bins=np.arange(0,14))
```



Rooted in *algorithms*

- ✗ Approximate solutions
- ✓ Often convincing
- ✓ Non-trivial problems potentially captured cleanly with code

Repeat many times:

*Flip 100 coins*

Record *the number of heads*

Analyze outcomes for all trials

# How to calculate an event's probability?

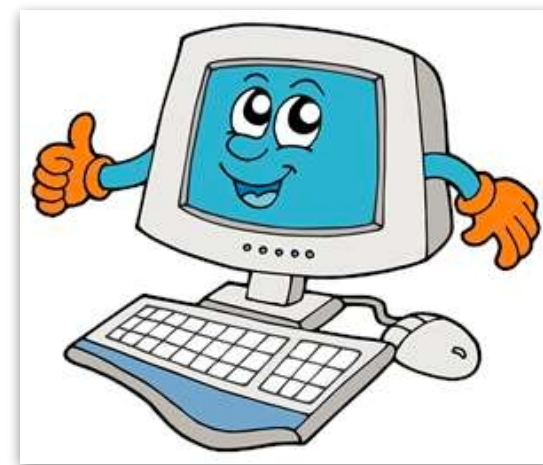
## Computers (simulation)

```
N = 1000000 #Roll the dice 1 million times
option_a = np.random.choice(dice, N) + np.random.c
option_b = 2 * np.random.choice(dice, N)

print("Option A Mean: ", np.mean(option_a))
print("Option B Mean: ", np.mean(option_b))

Option A Mean: 7.003198
Option B Mean: 6.99884

samples = Table().with_columns("Option A", option_
samples.hist("Option A", bins=np.arange(0,14))
samples.hist("Option B", bins=np.arange(0,14))
```



Rooted in *algorithms*

- ✗ Approximate solutions
- ✓ Often convincing
- ✓ Non-trivial problems potentially captured cleanly with code

Repeat many times:

*Flip 100 coins*

Record *the number of heads*

Analyze *what proportion of the times the number of heads was between 40 and 60*

# Notebook: *Simulation*

# How to calculate an event's probability?

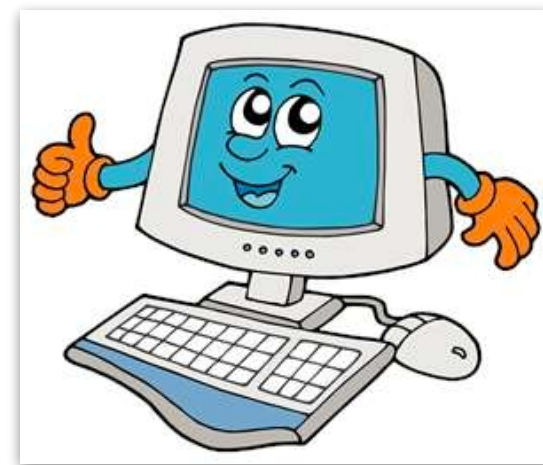
## Computers (simulation)

```
N = 1000000 #Roll the dice 1 million times
option_a = np.random.choice(dice, N) + np.random.c
option_b = 2 * np.random.choice(dice, N)

print("Option A Mean: ", np.mean(option_a))
print("Option B Mean: ", np.mean(option_b))

Option A Mean: 7.003198
Option B Mean: 6.99884

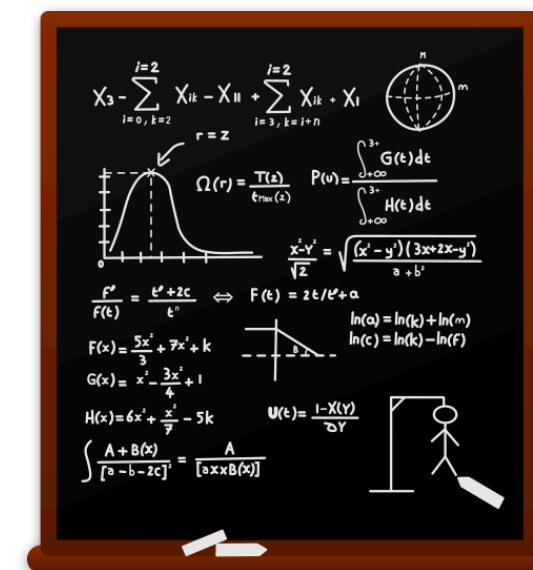
samples = Table().with_columns("Option A", option_
samples.hist("Option A", bins=np.arange(0,14))
samples.hist("Option B", bins=np.arange(0,14))
```



Rooted in *algorithms*

- ✗ Approximate solutions
- ✓ Often convincing
- ✓ Non-trivial problems potentially captured cleanly with code

## Math (analytical)



Rooted in *rules (axioms)*

- ✓ Exact solutions
- ✓ Straightforward for simple problems
- ✗ Non-trivial problems potentially difficult to analyze/verify

# Basic probability

If all outcomes are equally likely – for example, rolling a fair die or flipping a fair coin – then it's easy to compute a probability by counting:

$$P(A) = \frac{\text{number of outcomes that make } A \text{ happen}}{\text{total number of outcomes}}$$

How likely are you to get an even number when rolling a die?

$$\text{Even} = \{\text{□}, \text{□}, \text{□}\}$$

$$\text{All} = \{\text{□}, \text{□}, \text{□}, \text{□}, \text{□}, \text{□}\}$$

$$|\text{Even}| / |\text{All}| = 3 / 6 = 50\%$$

*In math, {...} denotes a set, and |...| denotes the number of elements.*

When there are two or more ways an event can happen, the *addition rule* says its probability is the sum of the probabilities for the different ways:

$$P(A) = P(\text{first way } A \text{ can happen}) + \\ P(\text{second way } A \text{ can happen}) + \\ \dots$$

For example, the probability of rolling a total of six when rolling two dice is:

$$\begin{aligned} P(\text{roll six}) = & P(\text{⊠}, \text{⊠}) + \\ & P(\text{⊠}, \text{⊠}) + \\ & P(\text{⊠}, \text{⊠}) + \\ & P(\text{⊠}, \text{⊠}) + \\ & P(\text{⊠}, \text{⊠}) \end{aligned}$$

The chance that event  $A$  will *not* occur:

$$P(\text{not } A) = 1 - P(A)$$

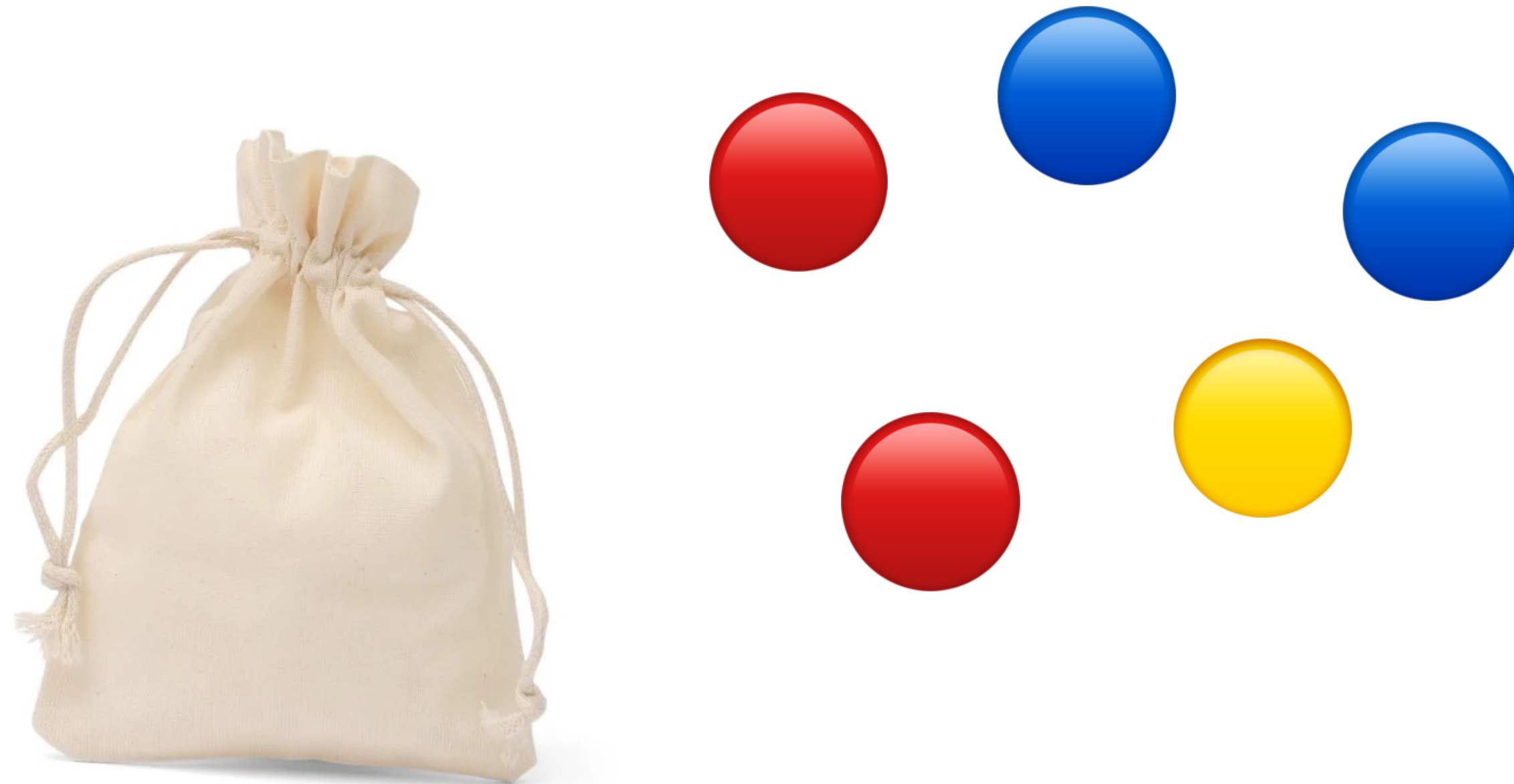
So, if  $P(\text{☁️🌧️}) = 0.7$  (that is, a 70% chance of rain)

then  $P(\text{not } \text{☁️🌧️}) = 1 - 0.7$   
 $= 0.3$  (that is, 30%)

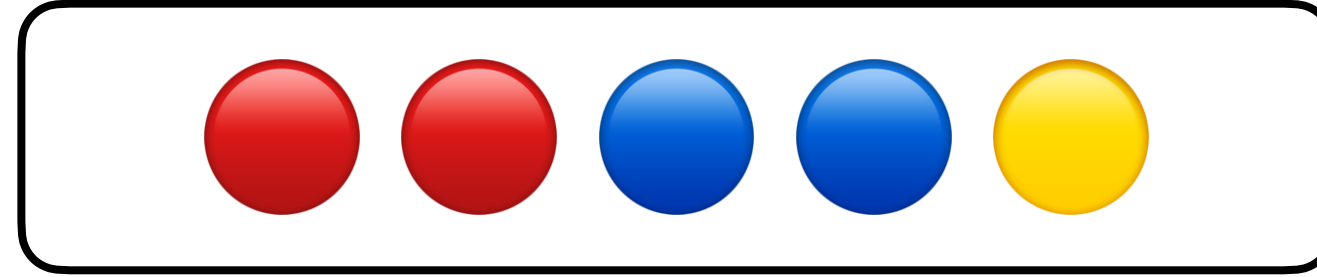
When you want to know the probability of events  $A$  and  $B$  *both* happening, you use the *multiplication rule*:

$$P(A \text{ and } B) = P(A) \cdot P(B \text{ given that } A \text{ happened})$$

For example, if you are taking colored marbles out of a bag, what is the probability of drawing a yellow marble *then* a blue marble?

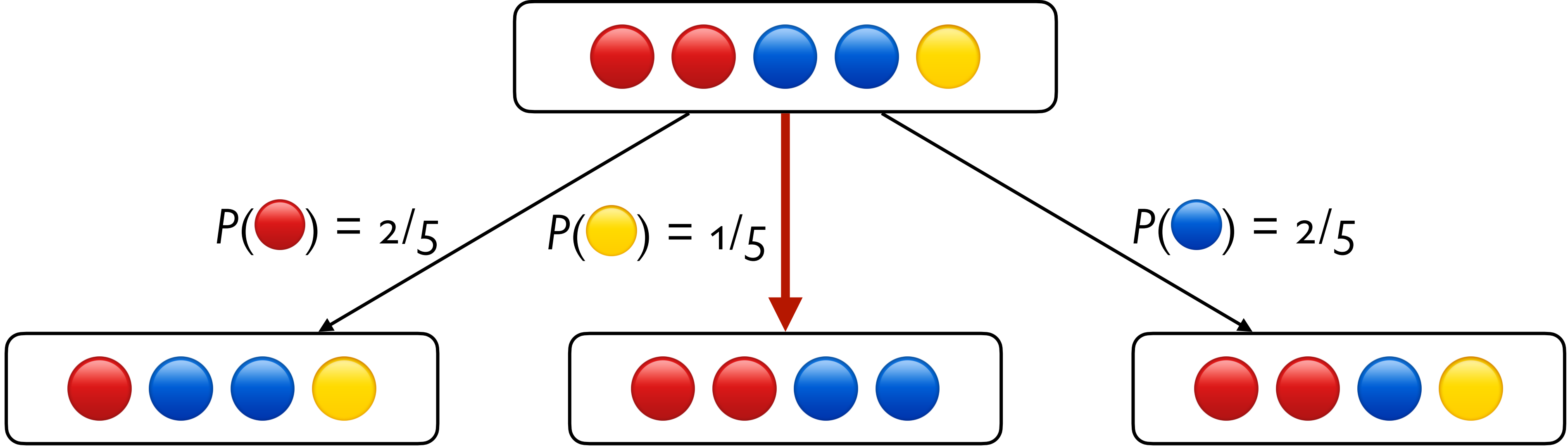


$P(\text{draw a } \text{yellow} \text{ then a } \text{blue}) = ?$



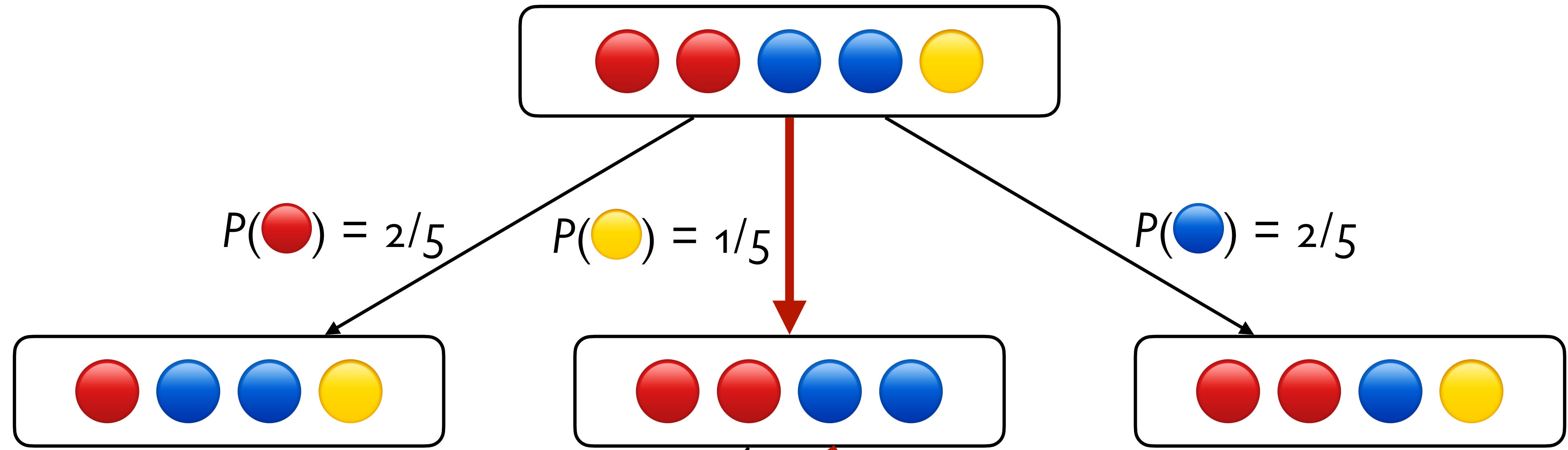
$P(\text{draw a } \text{yellow} \text{ then a } \text{blue}) = ?$

*First draw*

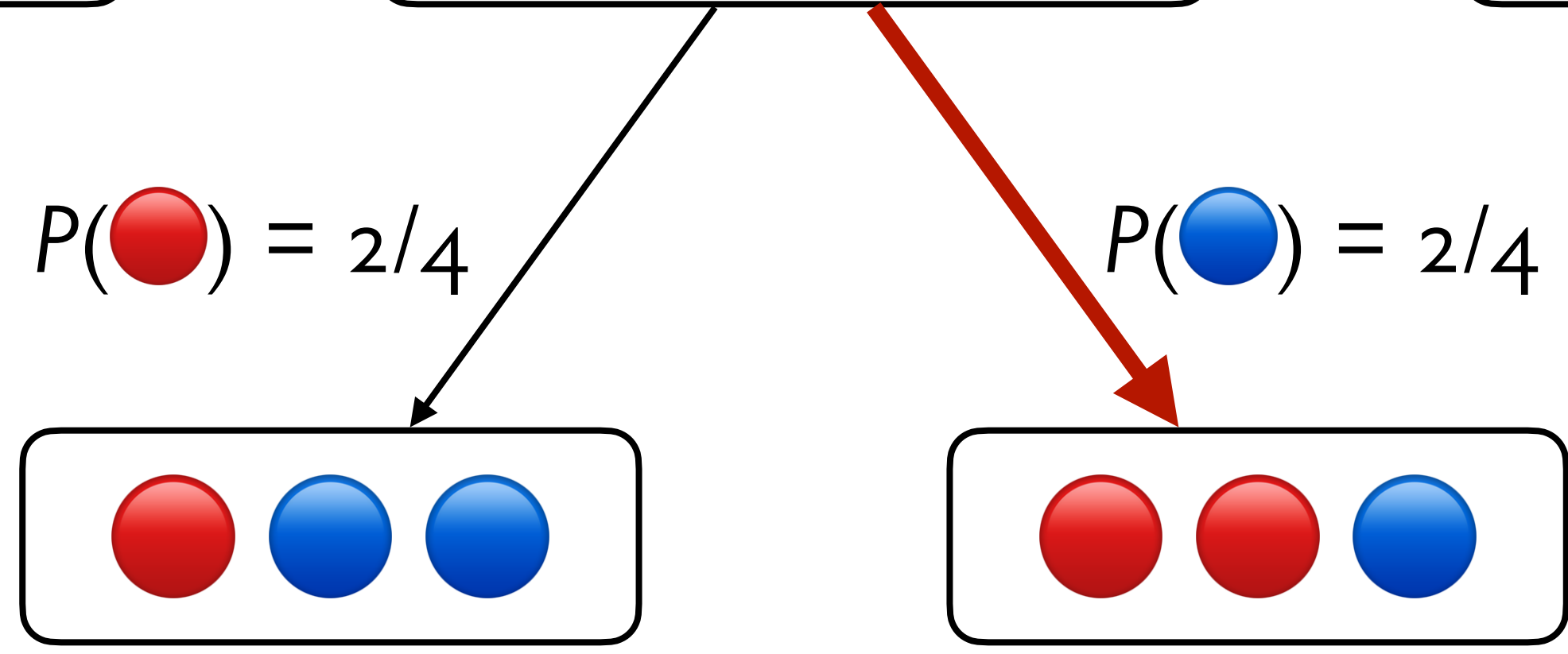


$P(\text{draw a } \text{yellow} \text{ then a } \text{blue}) = ?$

*First draw*

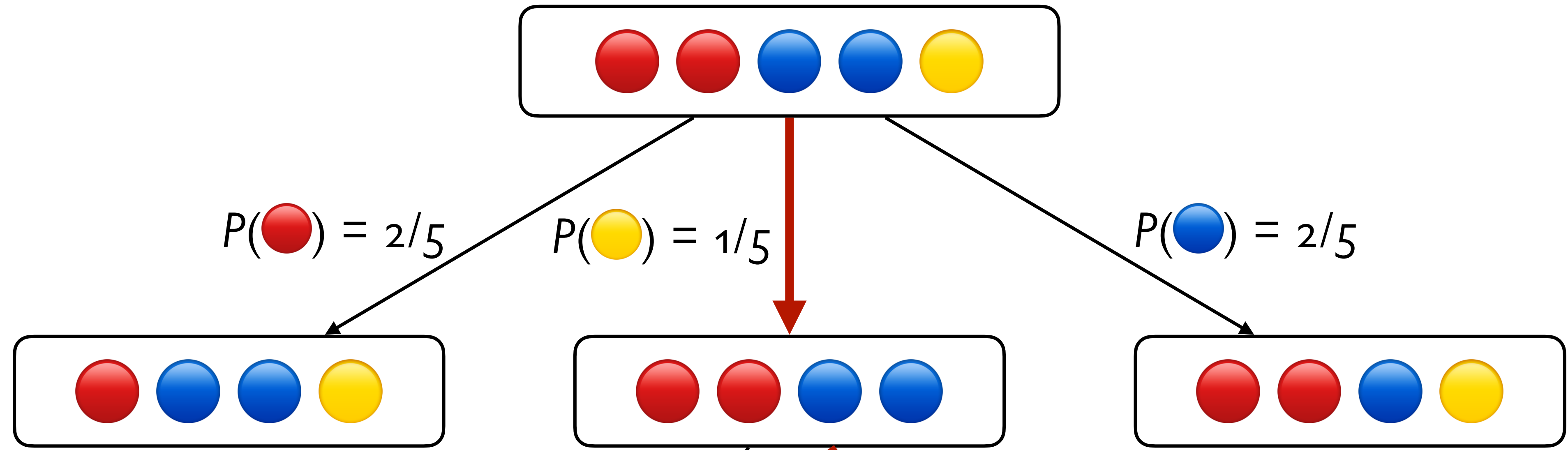


*Second draw*

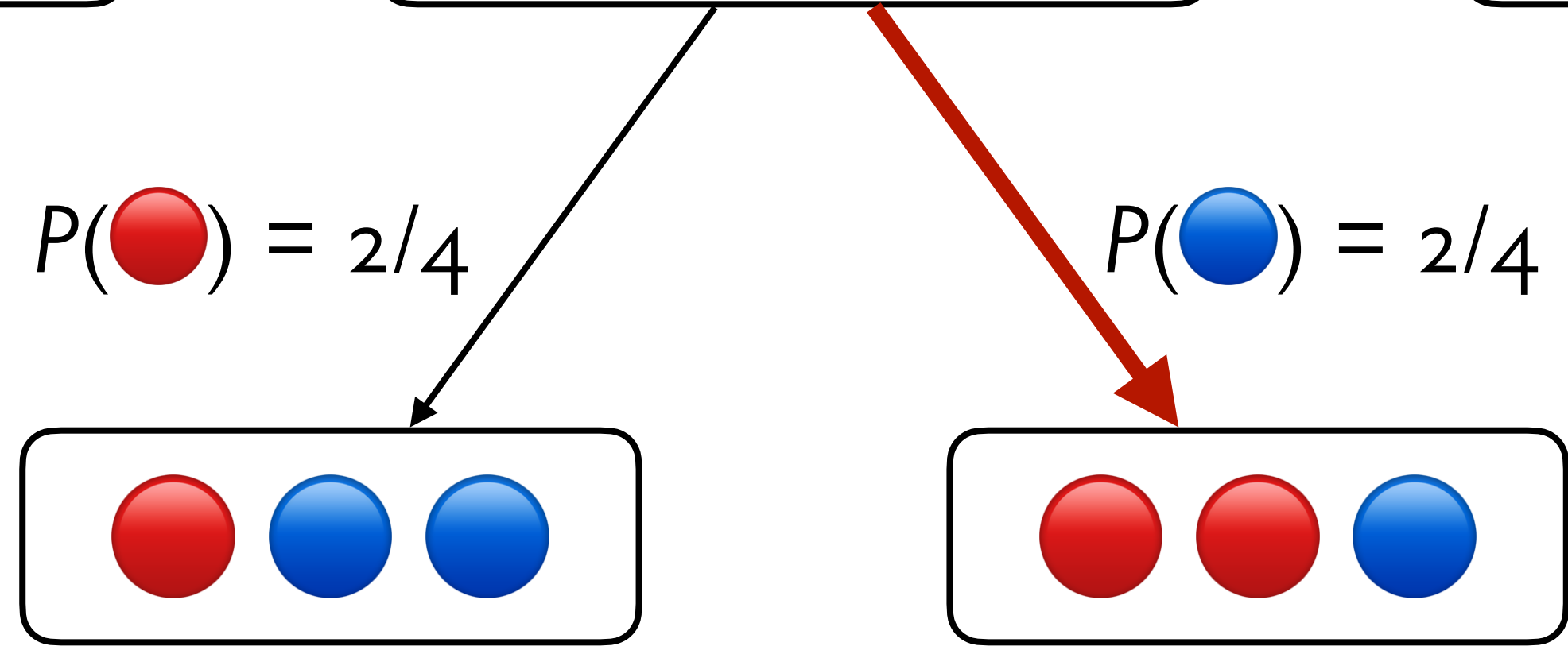


$$P(\text{draw a } \text{yellow} \text{ then a } \text{blue}) = \boxed{1/5 \cdot 2/4 = 10\%}$$

*First draw*



*Second draw*



# Practice

*Probability the sum of two dice is 12?*

# Practice

*Probability the sum of two dice is 12?*

$$P(\text{sum is 12}) = P(\text{roll } \text{⊞}) \cdot P(\text{roll } \text{⊞})$$

# Practice

*Probability the sum of two dice is 12?*

$$\begin{aligned} P(\text{sum is } 12) &= P(\text{roll } \text{⊞}) \cdot P(\text{roll } \text{⊞}) \\ &= 1/6 \cdot 1/6 \end{aligned}$$

# Practice

*Probability the sum of two dice is 12?*

$$\begin{aligned}P(\text{sum is } 12) &= P(\text{roll } \text{⊞}) \cdot P(\text{roll } \text{⊞}) \\ &= 1/6 \cdot 1/6 \\ &= 1/36\end{aligned}$$

# Practice



*Probability of all 4s in five rolls?*

# Practice



*Probability of all 4s in five rolls?*

$$P(\text{five } \begin{array}{|c|c|} \hline \cdot & \cdot \\ \hline \cdot & \cdot \\ \hline \end{array} \text{s}) = P(\text{roll } \begin{array}{|c|c|} \hline \cdot & \cdot \\ \hline \cdot & \cdot \\ \hline \end{array})^5$$

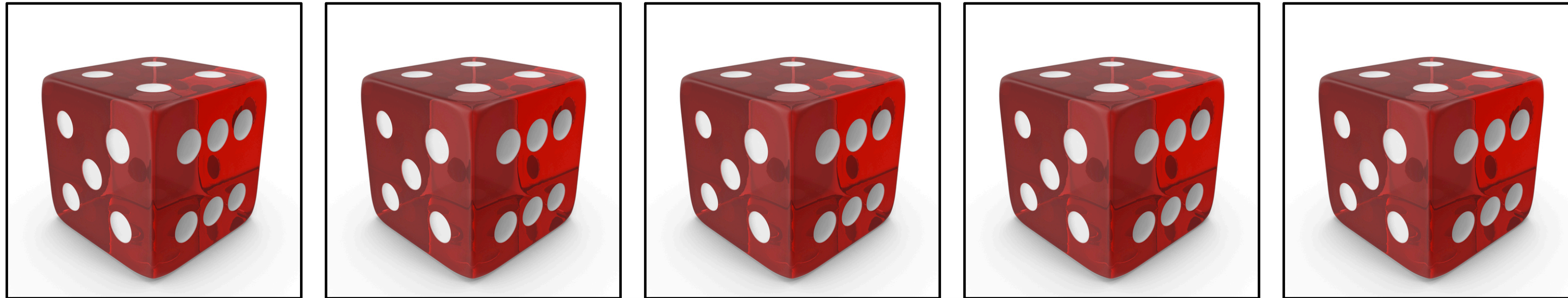
# Practice



*Probability of all 4s in five rolls?*

$$\begin{aligned} P(\text{five } \text{Ⓜ}\text{s}) &= P(\text{roll } \text{Ⓜ})^5 \\ &= (1/6)^5 \end{aligned}$$

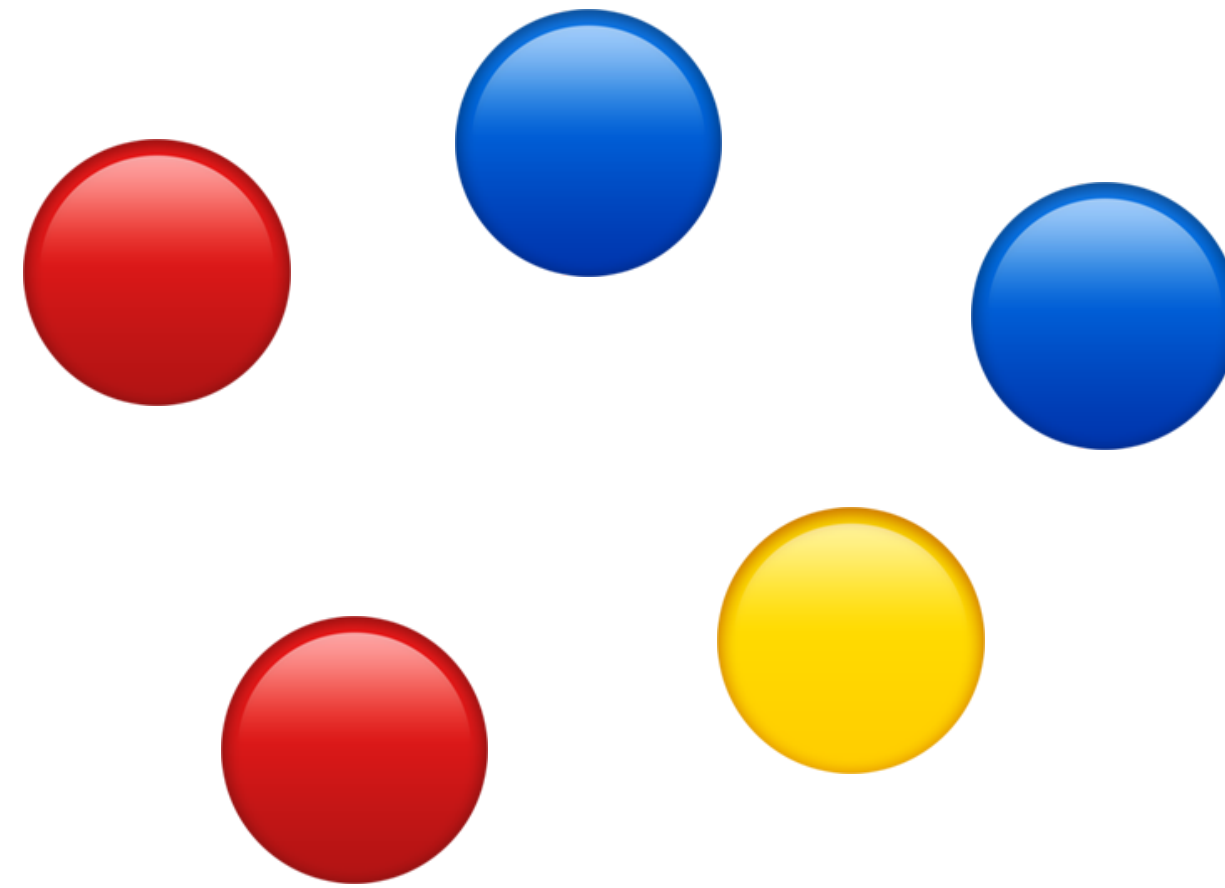
# Practice



*Probability of all 4s in five rolls?*

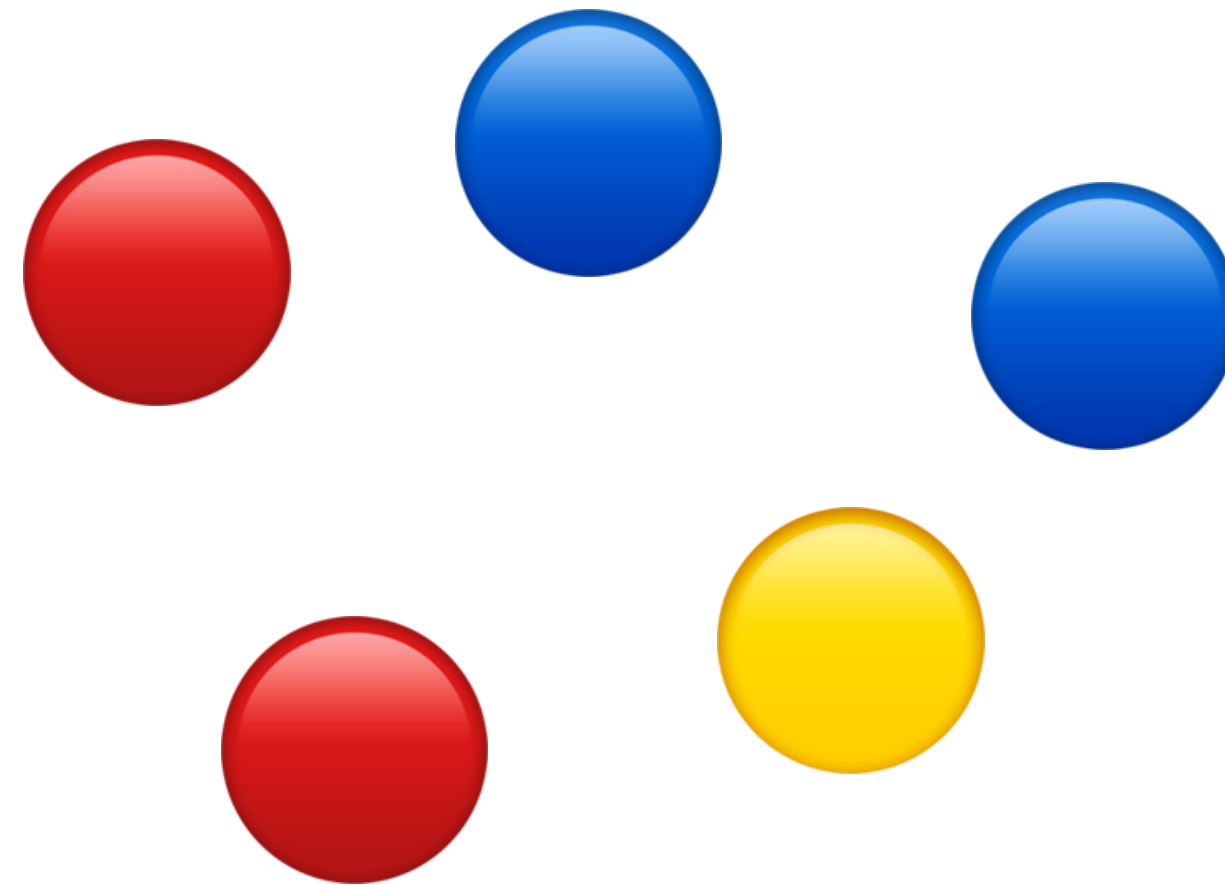
$$\begin{aligned}P(\text{five } \textcircled{4}\text{s}) &= P(\text{roll } \textcircled{4})^5 \\ &= (1/6)^5 \\ &= 0.00013\end{aligned}$$

# Practice



*Probability of drawing two of the same color?*

# Practice



*Probability of drawing two of the same color?*

$$\begin{aligned} P(\text{two of same color}) &= P(\text{blue then blue}) + P(\text{red then red}) + P(\text{yellow then yellow}) \\ &= \frac{2}{5} \cdot \frac{1}{4} + \frac{2}{5} \cdot \frac{1}{4} + \frac{1}{5} \cdot 0 \\ &= \frac{2}{20} + \frac{2}{20} + 0 \\ &= \frac{1}{5} \end{aligned}$$

# Probability and distributions

If you have a random quantity with various possible values, then its *probability (exact) distribution* associates

all the possible values of the quantity  
with the probability of each of those values.



*Sum of two dice rolls*

If you have a random quantity with various possible values, then its *probability (exact) distribution* associates

all the possible values of the quantity  
with the probability of each of those values.



*Sum of two dice rolls*

*Random quantity*

If you have a random quantity with various possible values, then its *probability (exact) distribution* associates

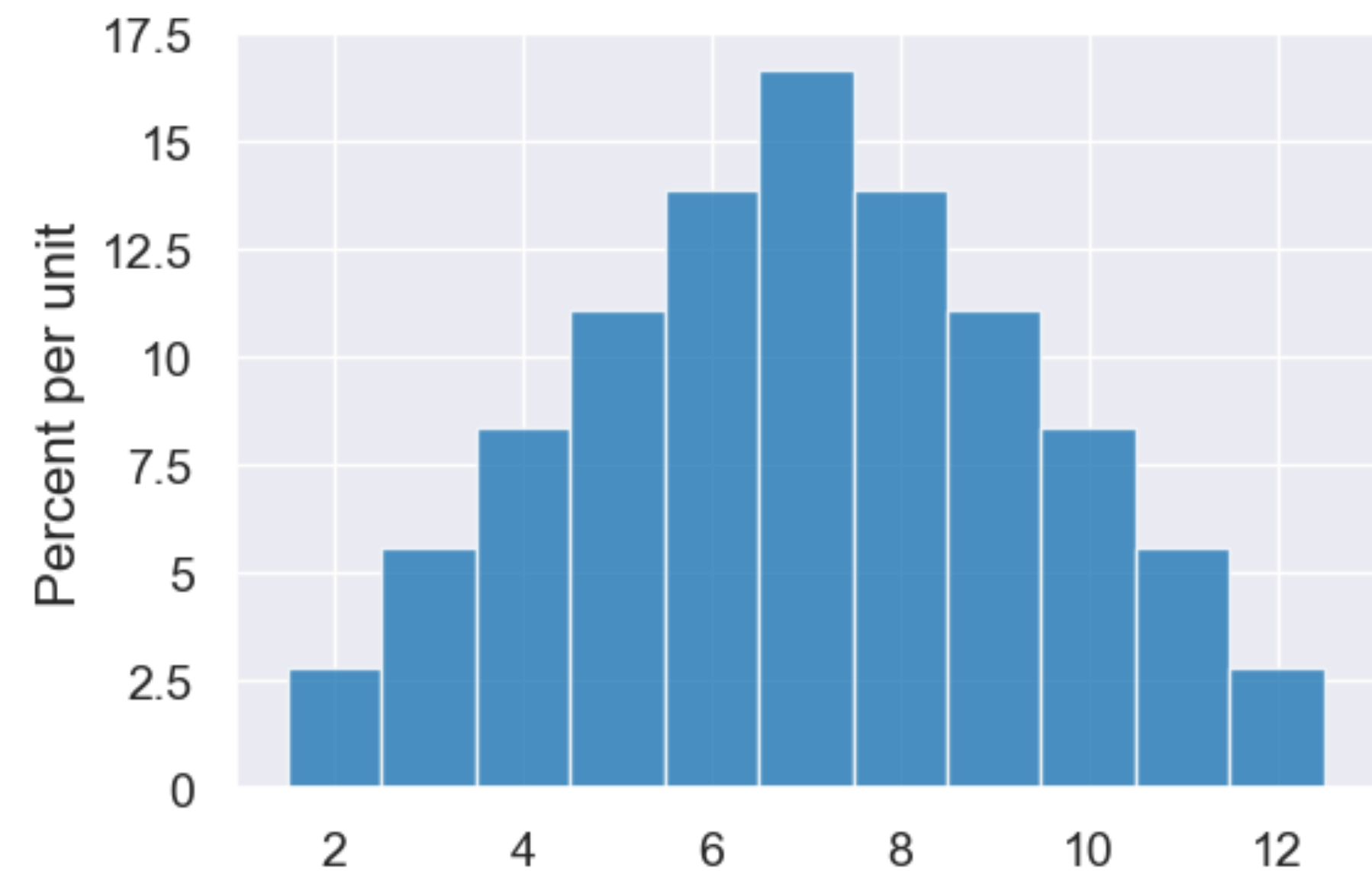
all the possible values of the quantity

with the probability of each of those values.



*Sum of two dice rolls*

*Random quantity*



*Every possible outcome*

If you can do the math, you can work out the probability distribution without ever simulating it:



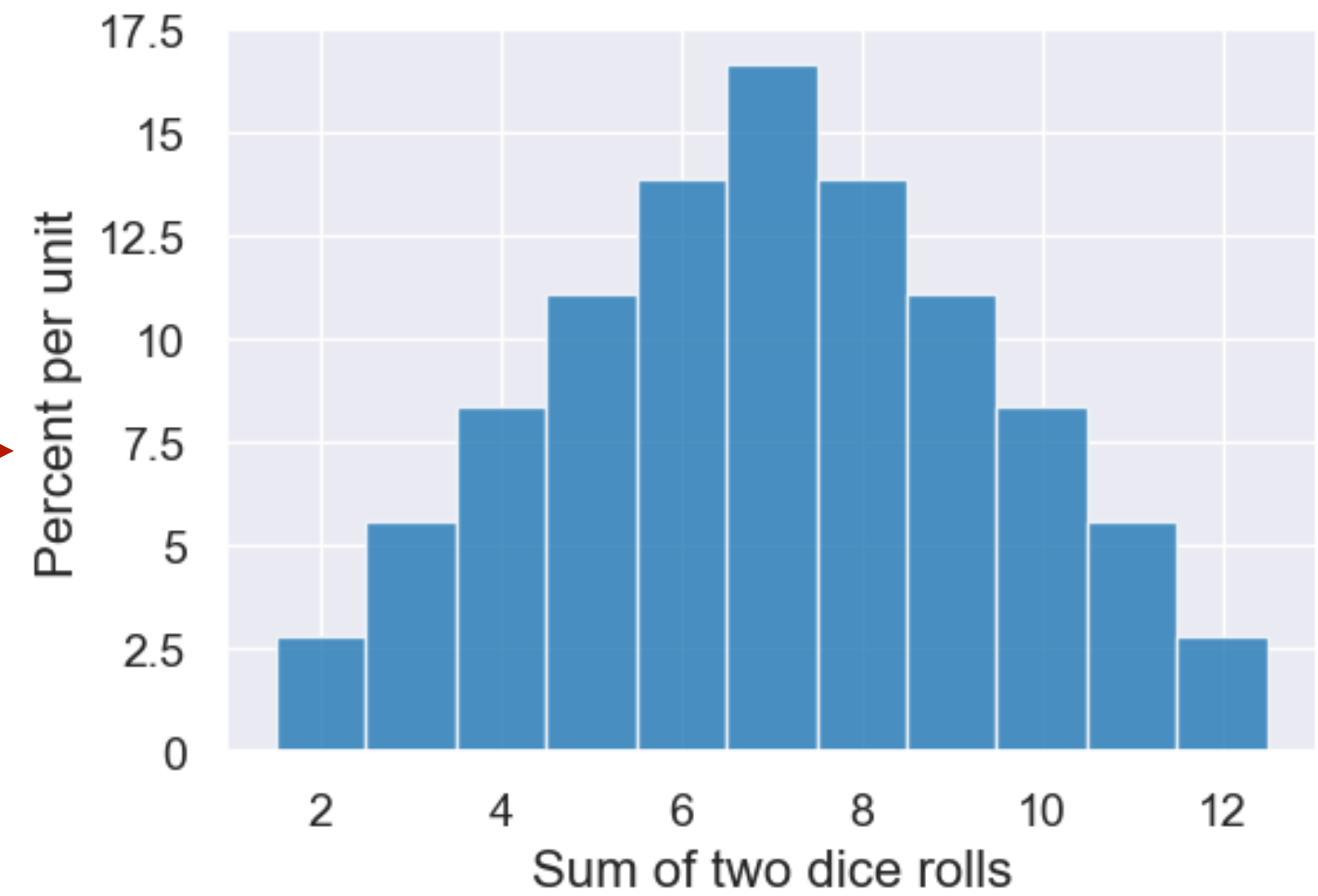
*Sum of two dice rolls*



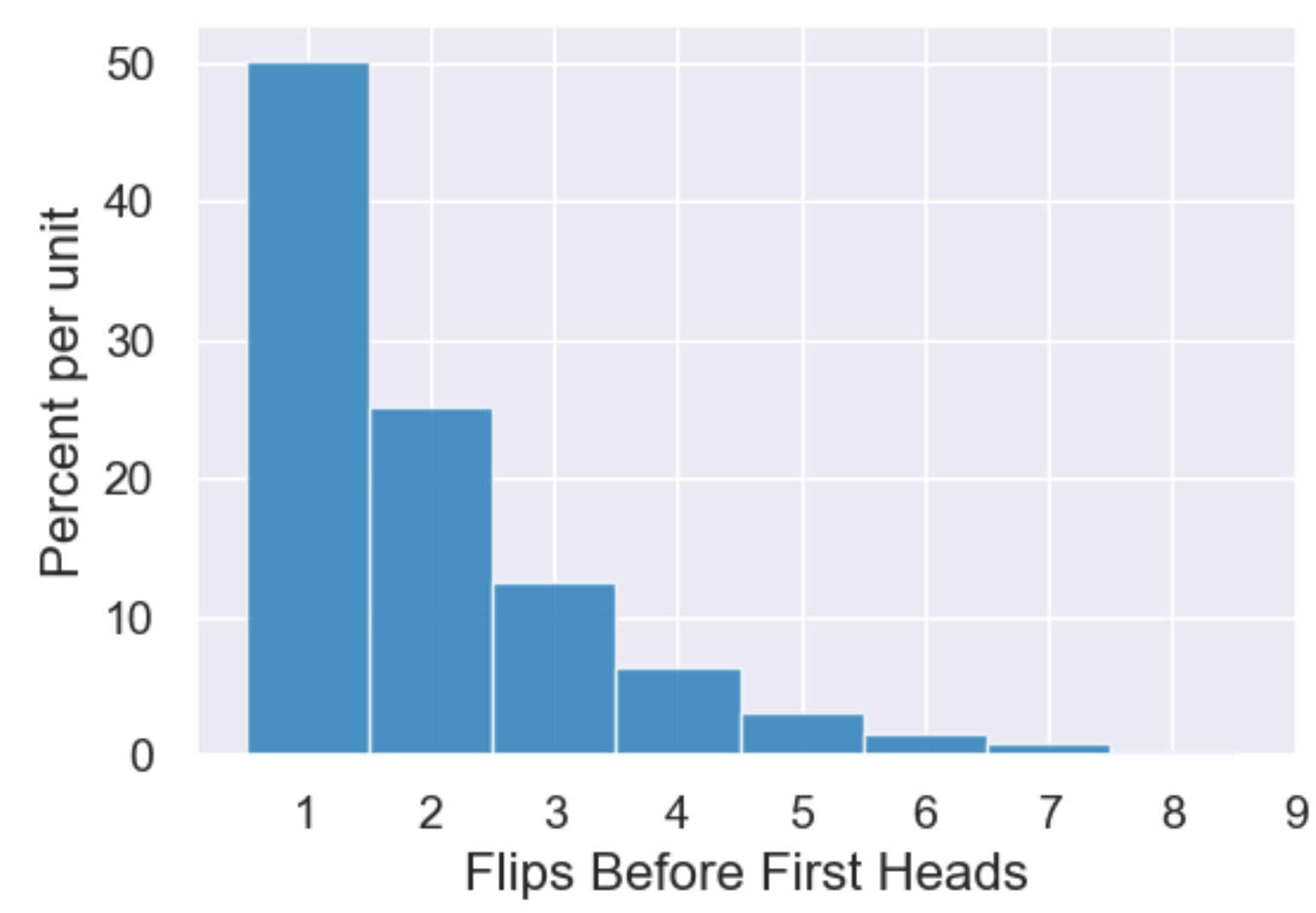
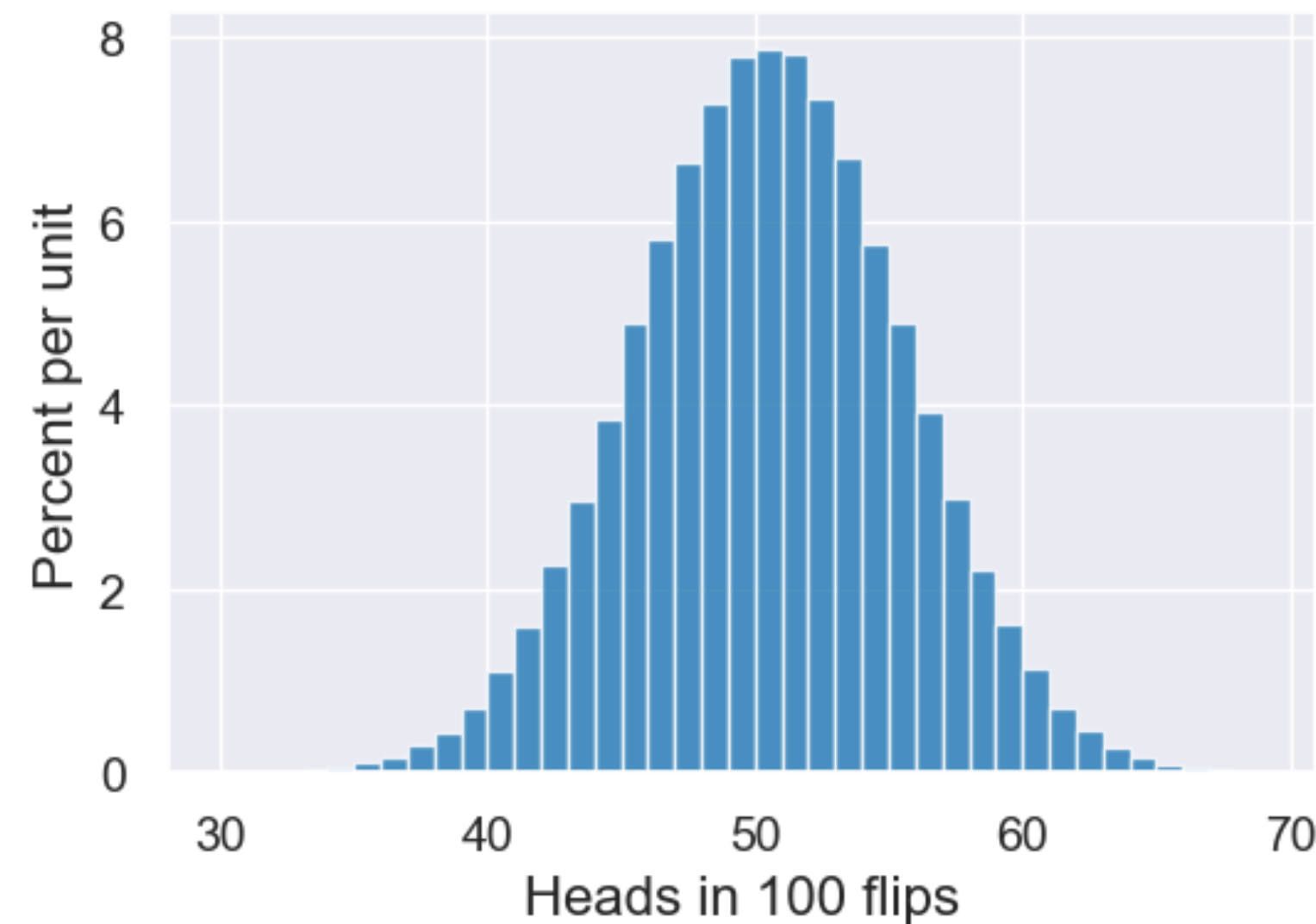
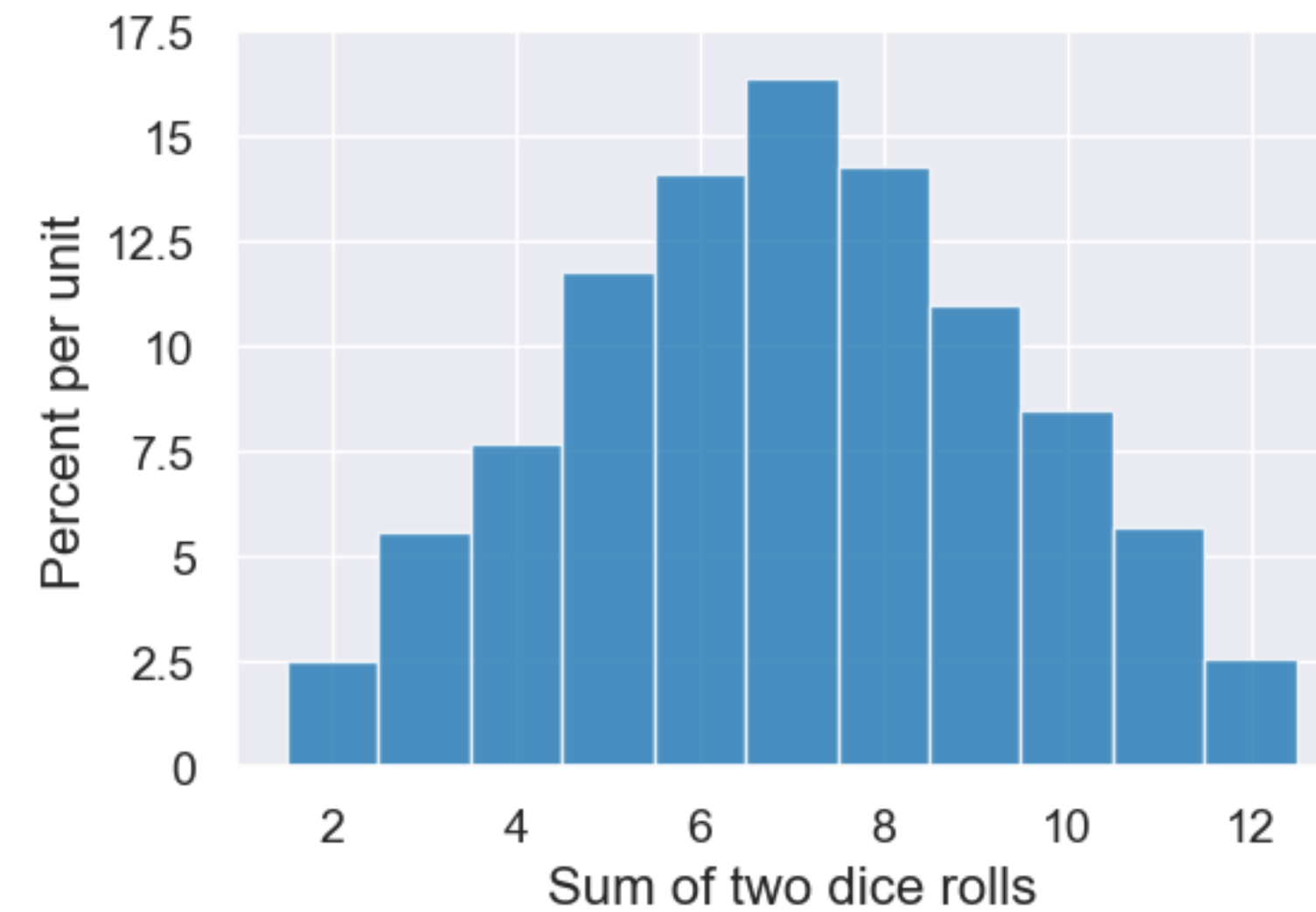
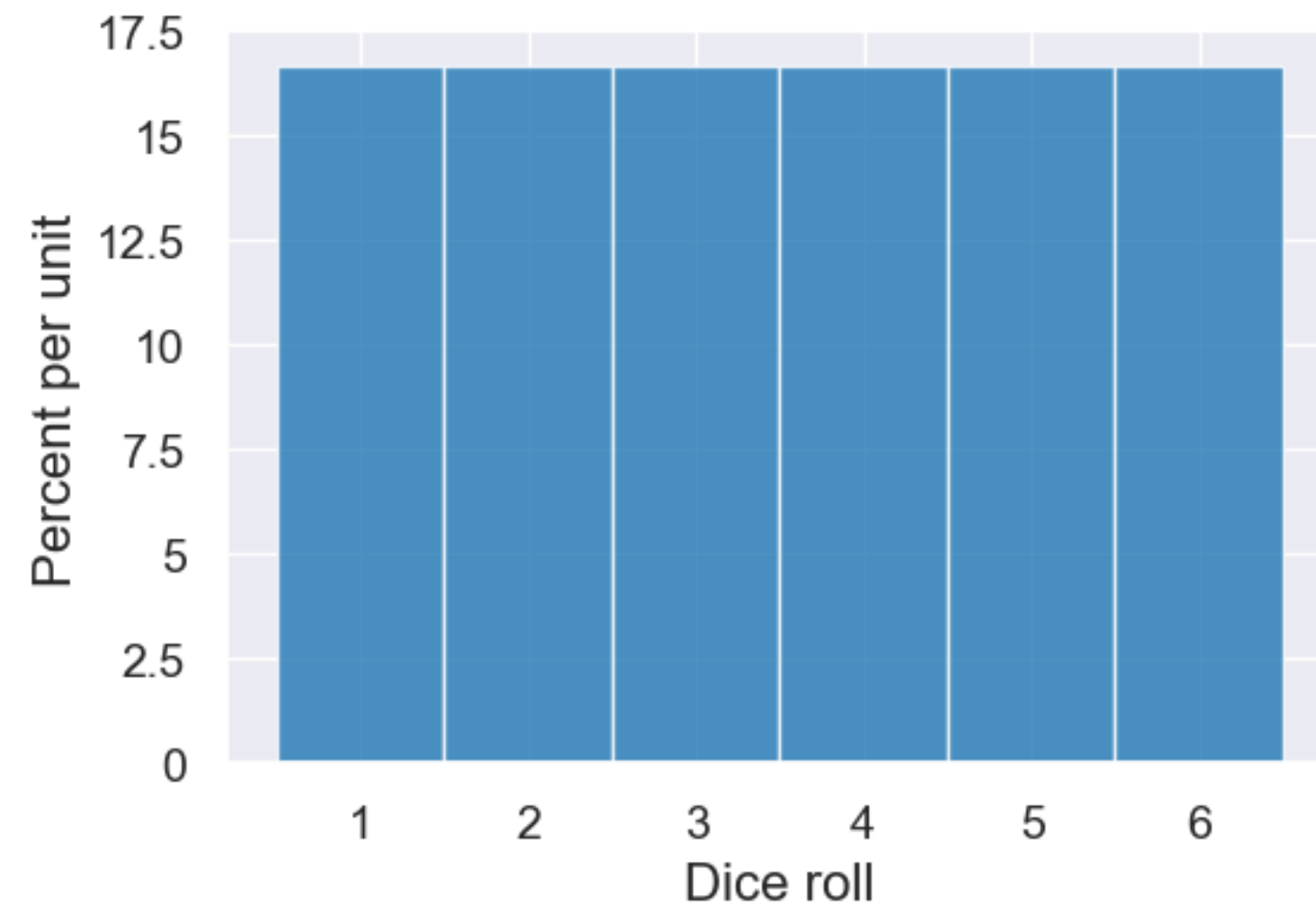
*Employ probability rules*

$$P(\text{sum is } 2) = P(\square, \square)$$
$$= 1/6 \cdot 1/6$$
$$P(\text{sum is } 3) = P(\square, \square) +$$
$$P(\square, \square)$$
$$= \dots$$

...

$$P(\text{sum is } 12) = \dots$$


# Probability distributions for other random quantities



# How to calculate an event's probability?

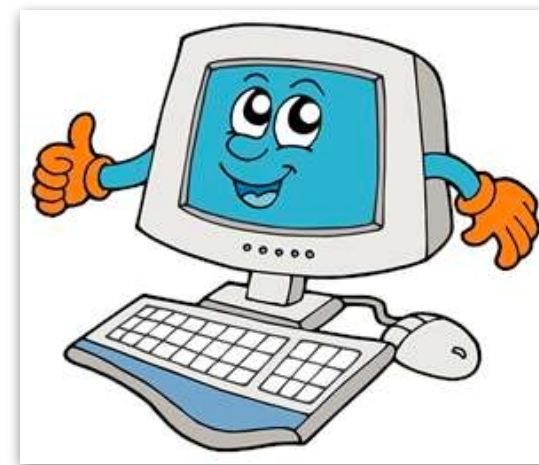
## Computers (simulation)

```
N = 1000000 #Roll the dice 1 million times
option_a = np.random.choice(dice, N) + np.random.c
option_b = 2 * np.random.choice(dice, N)

print("Option A Mean: ", np.mean(option_a))
print("Option B Mean: ", np.mean(option_b))

Option A Mean: 7.003198
Option B Mean: 6.99884

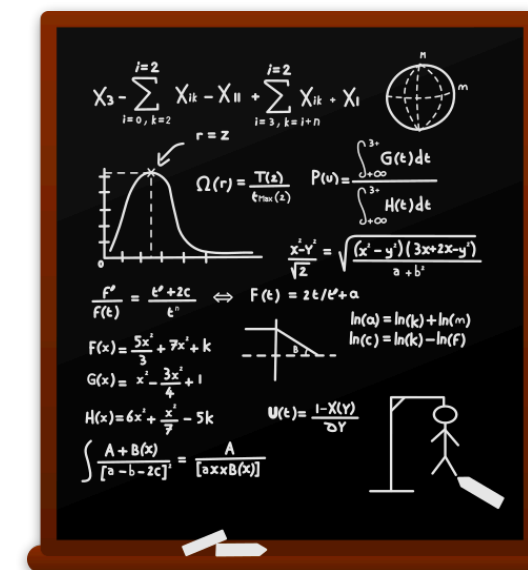
samples = Table().with_columns("Option A", option_
samples.hist("Option A", bins=np.arange(0,14))
samples.hist("Option B", bins=np.arange(0,14))
```



Rooted in *algorithms*

- ✗ Approximate solutions
- ✓ Often convincing
- ✓ Non-trivial problems potentially captured cleanly with code

## Math (analytical)

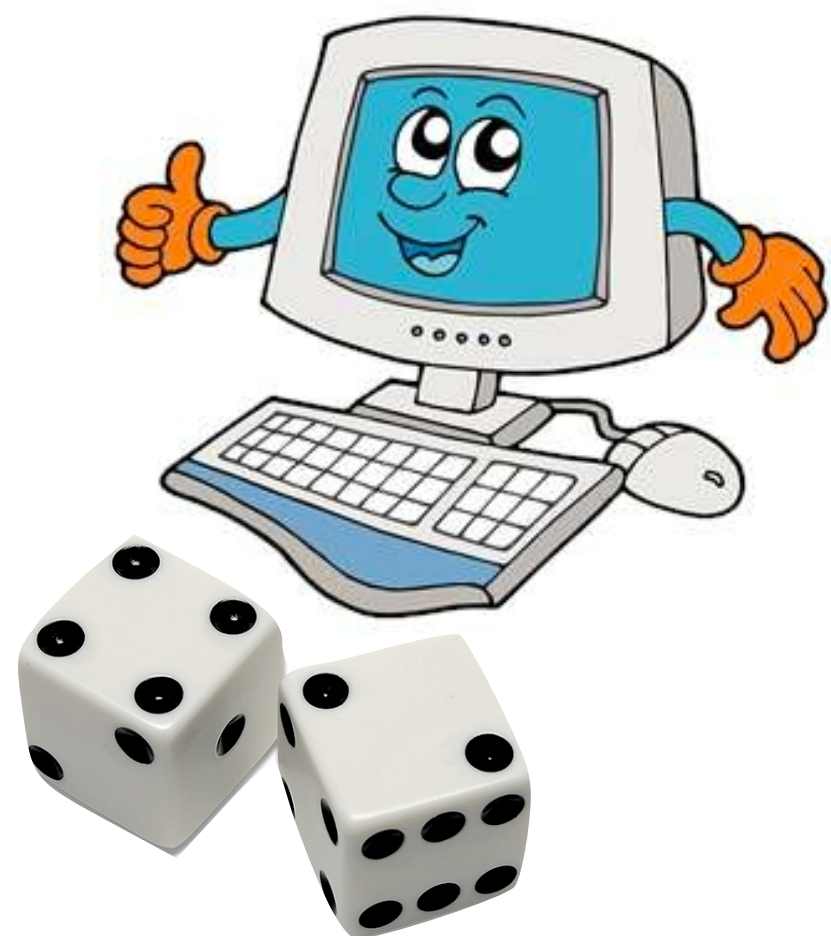


Rooted in *rules (axioms)*

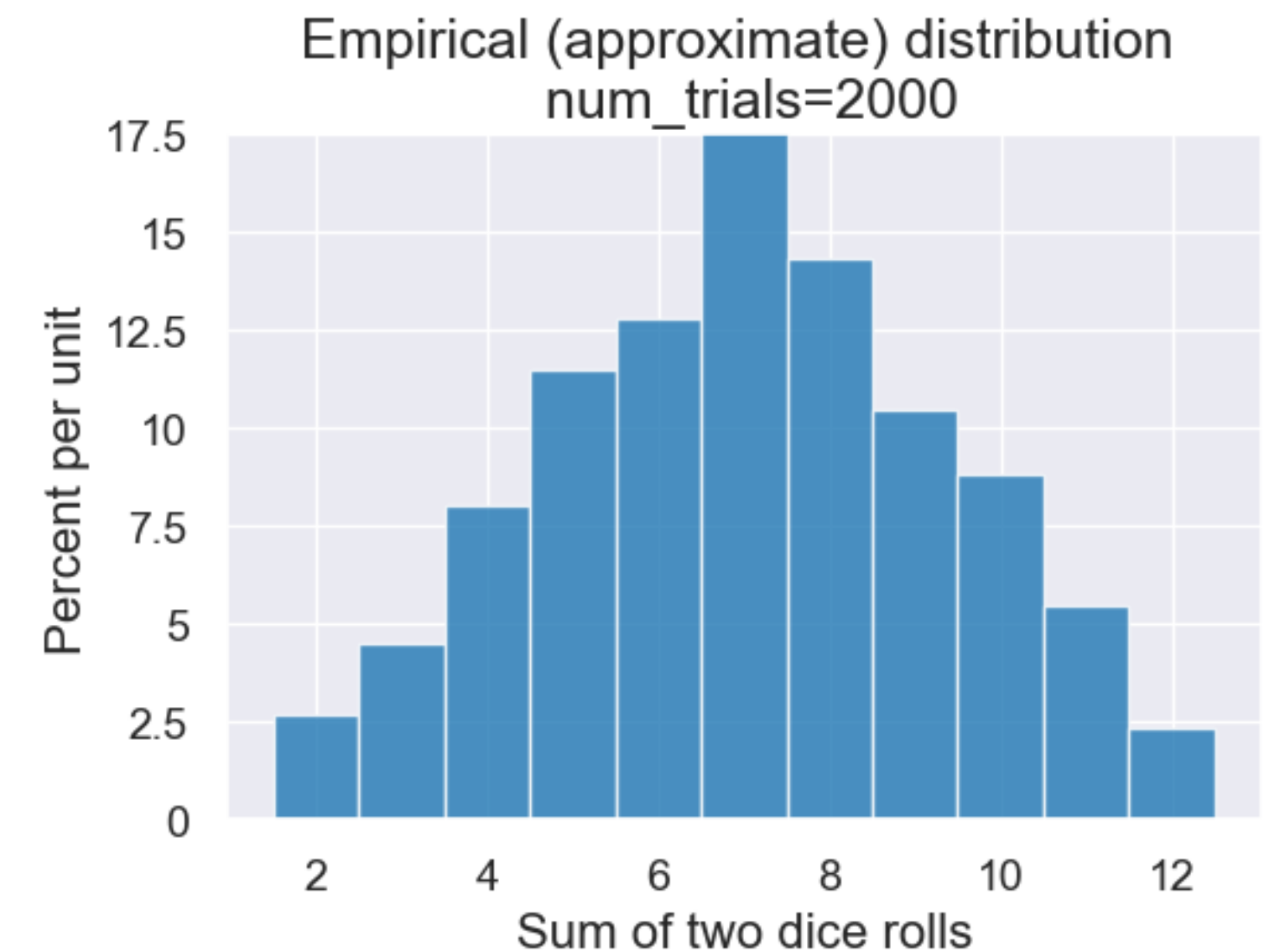
- ✓ Exact solutions
- ✓ Straightforward for simple problems
- ✗ Non-trivial problems potentially difficult to analyze/verify

An *empirical (approximate) distribution* consists of observations, which can be from repetitions of an experiment. It associates

all the unique values you actually *observed* with the proportion of times each value appeared.



Outcomes: 9, 6, 8, 9, 5, 8, 8, 8, 8,  
6, 9, 3, 3, 6, 9, 10, 6, 10, 7, 6, ...

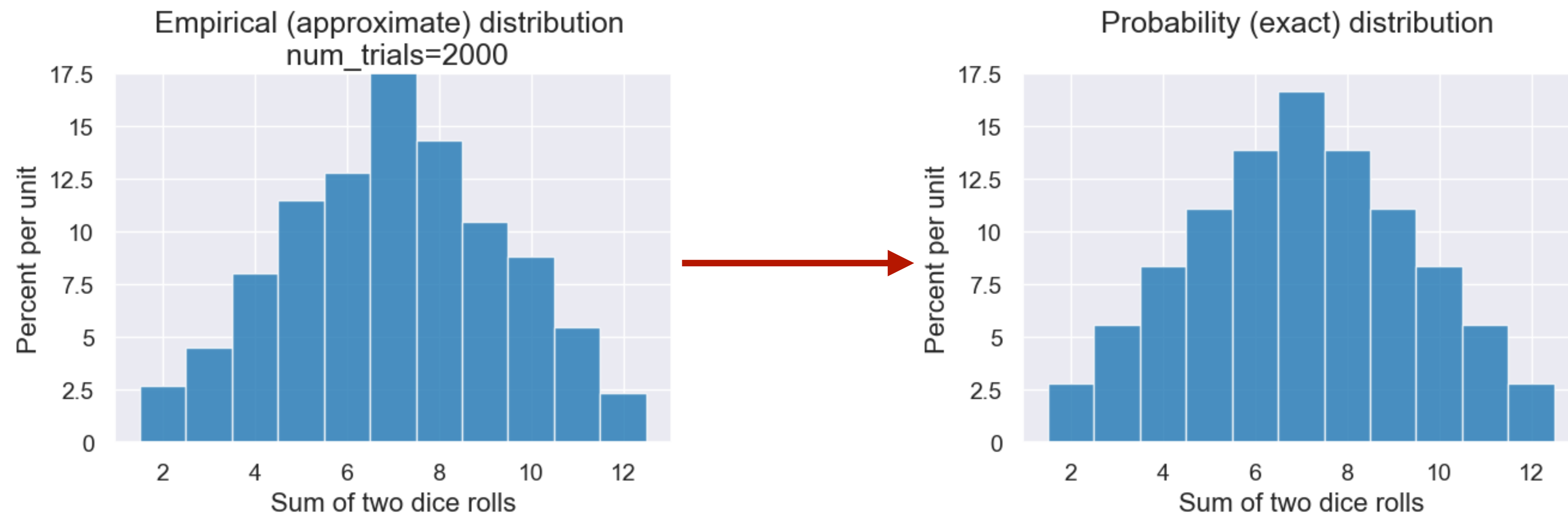


Notebook:

*Empirical distribution: Sum of two dice*

The *law of averages* (or *law of large numbers*):

If a chance experiment is repeated many times, independently and under the same conditions, then the proportion of times that an event occurs *gets closer* to the theoretical probability of the event.



# Empirical distribution of a sample

If the sample size is large,

then the empirical distribution of a uniform random sample resembles the distribution of the population,

with high probability.

# Real-world distributions and sampling

We could only simulate rolling two dice and taking their sum because we knew the true likelihood of each outcome for rolling a single die:

$$P(\text{1}) = 1/6$$

$$P(\text{2}) = 1/6$$

$$P(\text{3}) = 1/6$$

$$P(\text{4}) = 1/6$$

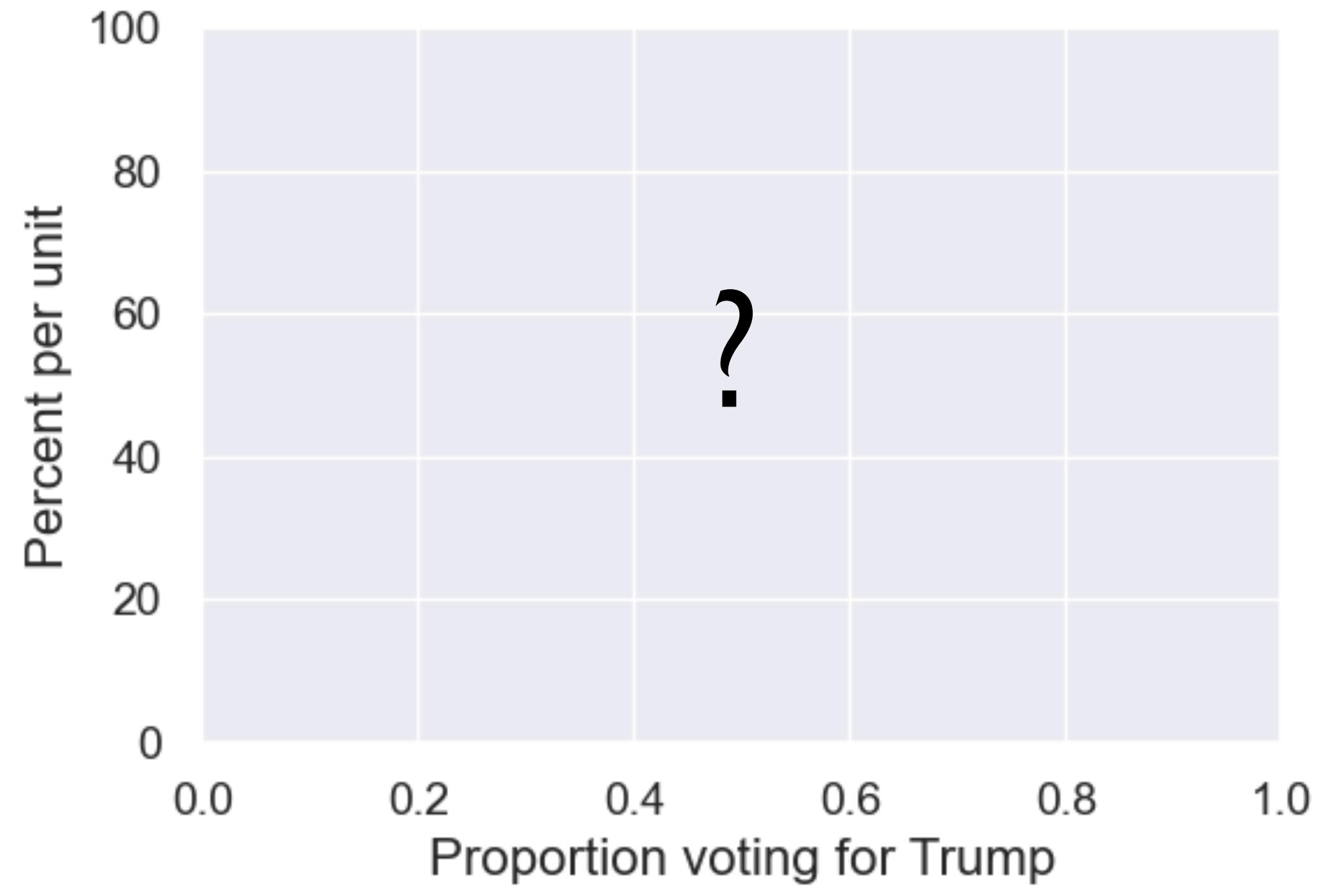
$$P(\text{5}) = 1/6$$

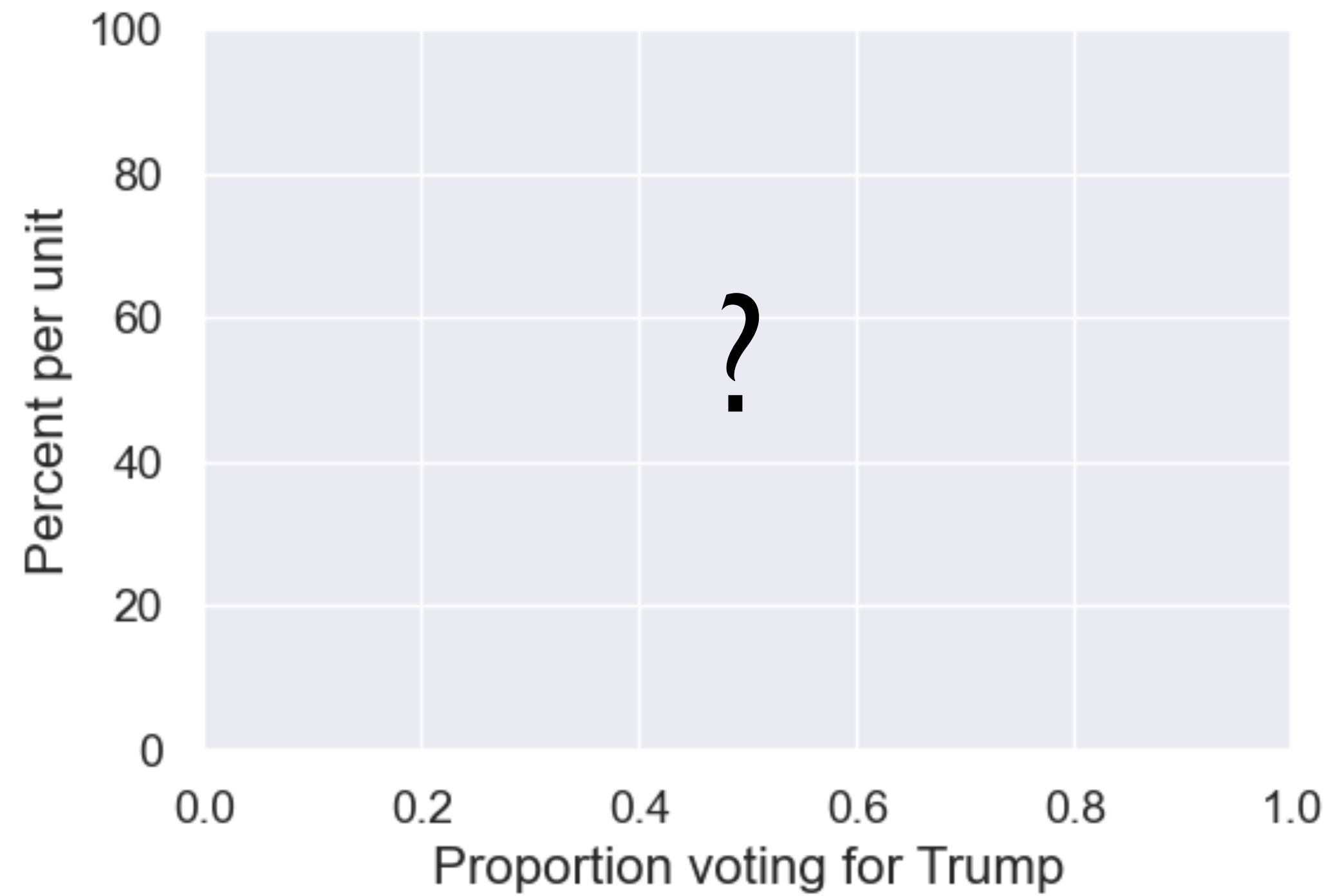
$$P(\text{6}) = 1/6$$



*Circa 2016*

*What is the probability distribution for a candidate's chance of getting some percent of votes in an upcoming election?*





*We don't know the true probability of whether each person will vote for a given candidate; we can't compute this distribution the way we did before!*



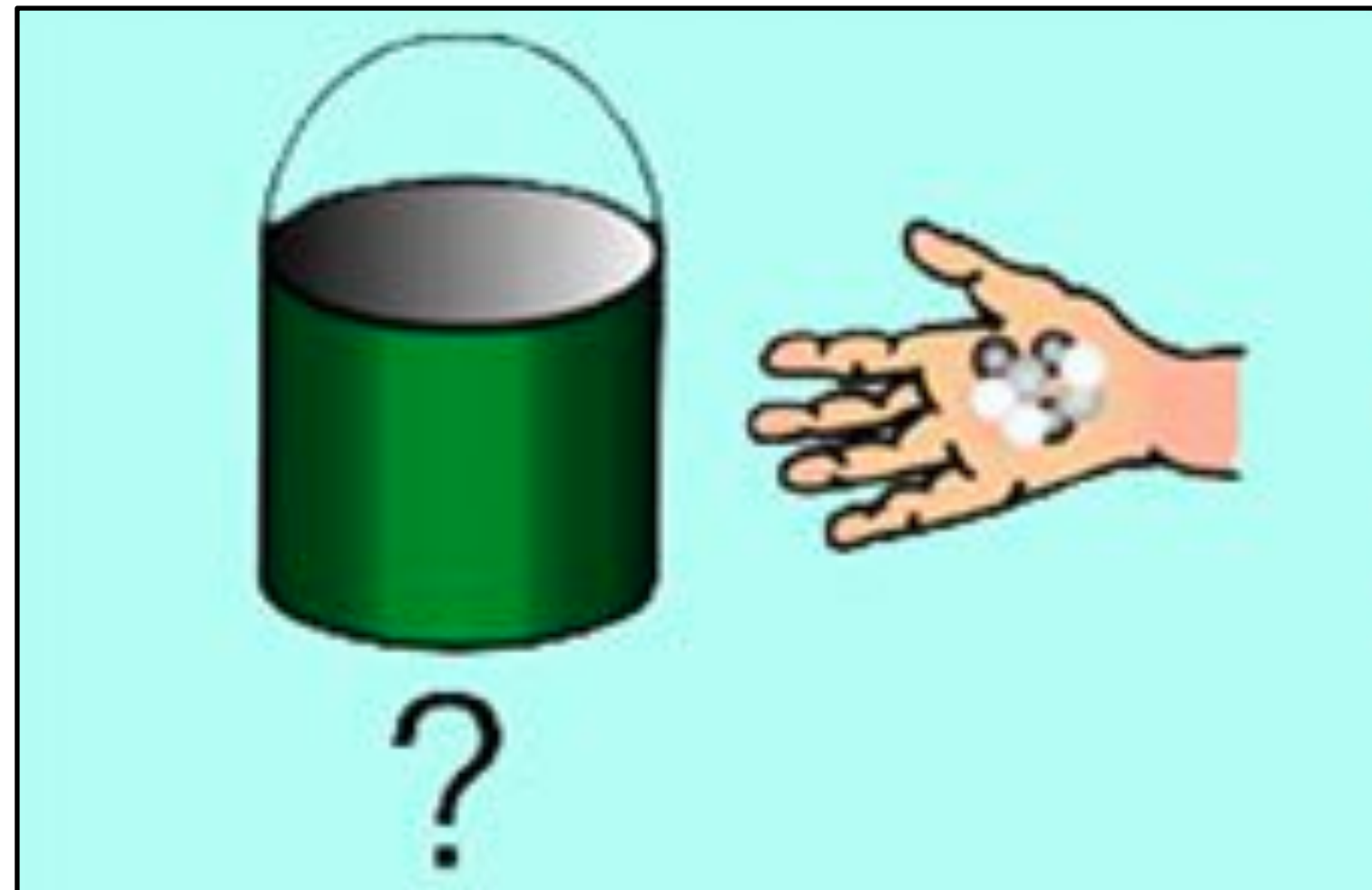
*Probability*

*Given the information in the pail, what is in your hand?*



*Probability*

*Given the information in the pail, what is in your hand?*



*Statistics*

*Given the information in your hand, what is in the pail?*

A *population* is a set of all elements from which a subset called a *sample* will be drawn.

How do we select our sample?

How do we draw meaningful conclusions using a sample?



*Statistics*

*Given the information in your hand, what is in the pail?*



Not all samples involve chance!

Here's an example of *deterministic sampling*:



*Pollster: Call every person with the first name "Bob" and ask him who he's voting for.*

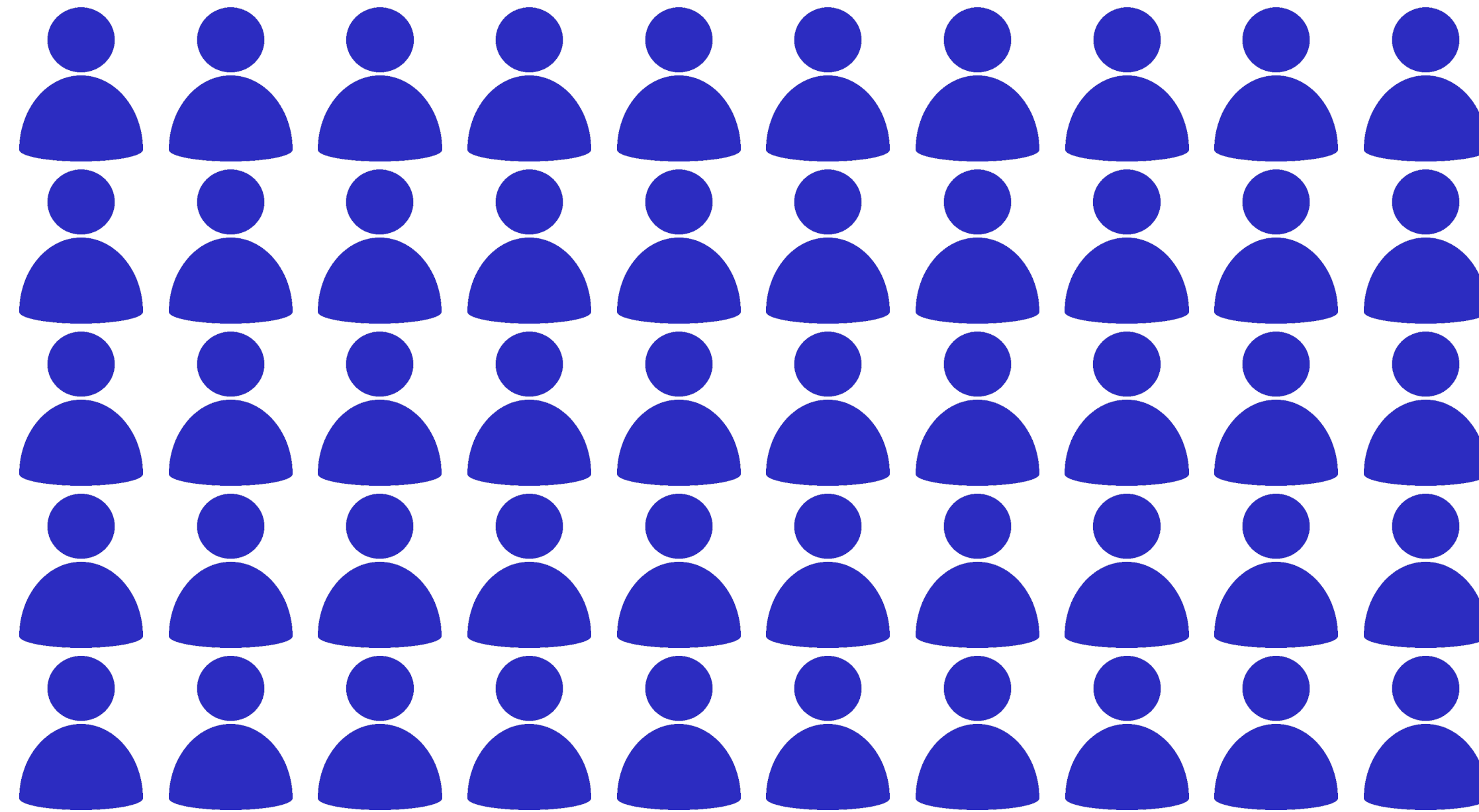
Not all samples involve chance!

Here's an example of *deterministic sampling*:



*Pollster: Call every person with the first name "Bob" and ask him who he's voting for.*

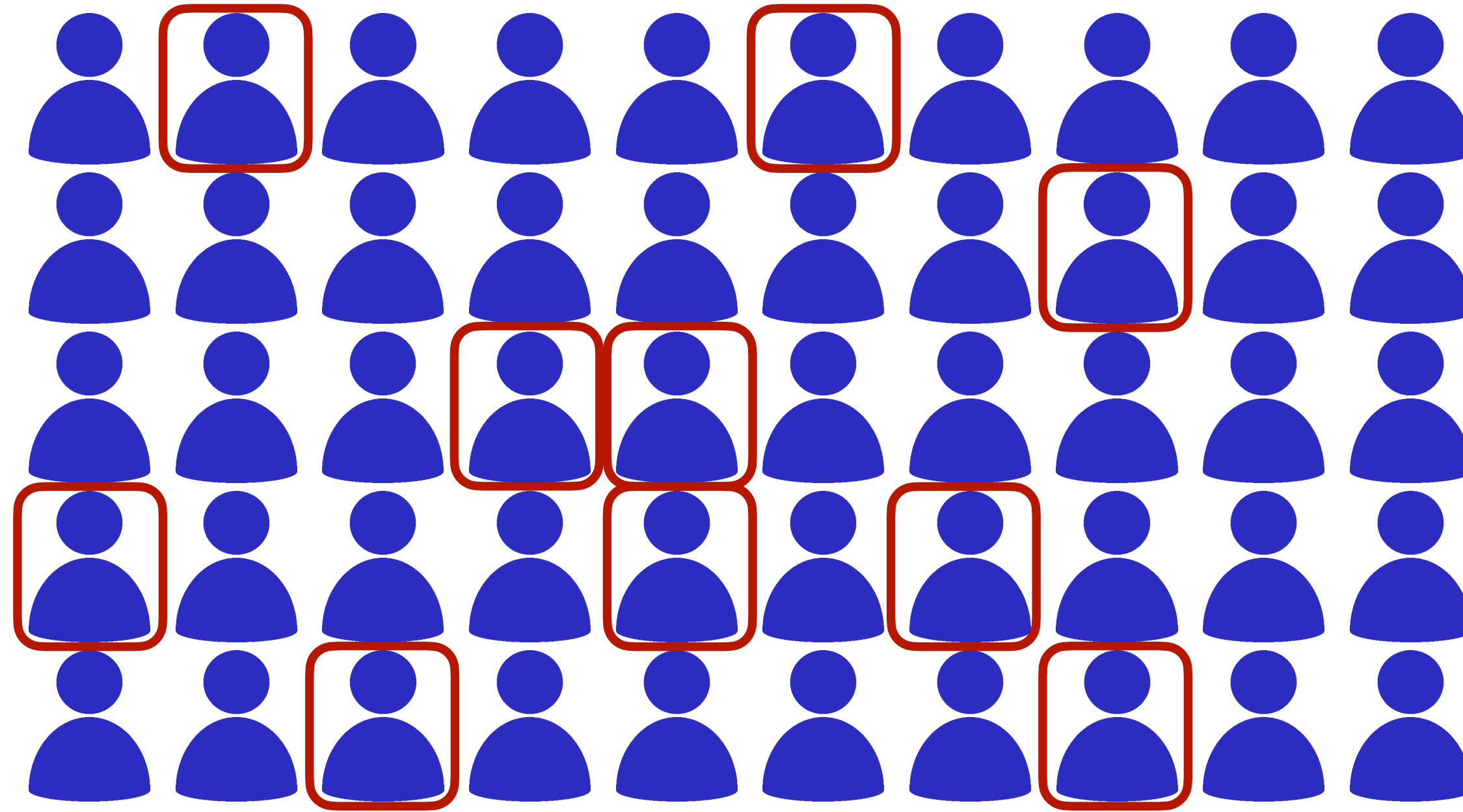
*What's wrong with this?*



In *random sampling*, each member of the population has some probability of being picked.



*Example:  
Random-digit dialing*



In *random sampling*, each member of the population has some probability of being picked.

HOME SEARCH The New York Times

Alaska Ariz. Colo. Fla. Ga. Ind. Iowa Miss. Mo. N.C. N.H. Nev. Ohio Pa. Wis.

# Who Will Be President?

By JOSH KATZ

The estimates on this page are based on pre-election polls. For an estimate including results, The Times is providing live forecasts on election night.

**President Senate House**

## Hillary Clinton has an 85% chance to win.

Last updated Tuesday, November 8 at 10:20 PM ET

CHANCE OF WINNING

85% Hillary Clinton 15% Donald J. Trump

Forecast history Recent changes State by state Other forecasts Likely scenarios Explore paths

The Upshot's elections model suggests that Hillary Clinton is favored to win the presidency, based on the latest state and national polls. A victory by Mr. Trump remains possible: Mrs. Clinton's chance of losing is about the same as the probability that an N.F.L. kicker misses a 37-yard field goal.

For months, we've been updating our estimates with each new poll. Today, it's Election Day, what we've all been waiting for, and there will be no more updates. You can chart different paths to victory below. Here's how our estimates have changed over time:

Month	Hillary Clinton (%)	Donald Trump (%)
June	60	40
July	75	25
August	80	20
September	80	20
October	80	20
November 8	85	15



"All the News That's Fit to Print" The New York Times Late Edition

VOL. CLXVI ... No. 57,411 ... © 2016 The New York Times Company NEW YORK, WEDNESDAY, NOVEMBER 9, 2016 \$2.50

# TRUMP TRIUMPHS

## OUTSIDER MOGUL CAPTURES THE PRESIDENCY, STUNNING CLINTON IN BATTLEGROUND STATES

Donald J. Trump voting on Tuesday at P.S. 59 in Manhattan. His defeat of Hillary Clinton defied late polls and was a repudiation of the establishment.

### WORKING CLASS SPEAKS

Blue-Collar Whites Gave Stinging Rebuke to Democratic Party

By PATRICK HEALY and JONATHAN MARTIN

Donald John Trump was elected the 45th president of the United States on Tuesday in a stunning culmination of an explosive, pragmatic and polarizing campaign that took relentless aim at the institutions and long-held ideals of American democracy.

The surprise outcome, defying late polls that showed Hillary Clinton with a modest but persistent edge, threatened cohesion throughout the country and the world, where skeptics had watched with alarm as Mr. Trump's unvarnished overtures to disillusioned voters took hold.

The triumph for Mr. Trump, 70, a real estate developer-turned-reality television star with no government experience, was a powerful rejection of the establishment forces that had assembled against him. From the world of business to government, and the consensus they had forged on everything from trade to immigration.

The results amounted to a repudiation, not only of Mrs. Clinton, but of President Obama, whose legacy is suddenly imperiled. And it was a decisive demonstration of power by a largely overlooked coalition of mostly blue-collar white and working-class voters who felt that the promises of the United States had slipped their grasp amid decades of globalization and multiculturalism.

In Mr. Trump, a birch-married Manhattanite who lives in a marble-wrapped, three-story penthouse apartment on Fifth Avenue, they found an improbable champion.

Mr. Trump's strong showing helped Republicans regain control of the Senate. Only one Republican-controlled seat, in Illinois, fell to Democrats early in the evening. And Senator Richard Burr of North Carolina, a Republican, easily won reelection in a race that had been among the country's most competitive. A handful of other Republican incumbents facing difficult races were running better than expected.

Continued on Election 2016, Page 5

### AMBRIDGE JOURNAL

#### A Blue-Collar Town in Decline And in Despair Turns to Trump

By TREP GABRIEL

AMBRIDGE, Pa. — As Donald J. Trump's surprisingly strong showing played out on television above Free's Diner bar, the men who by day carry pipes, hang dry wall and drive locomotives watched the returns with mounting distaste.

"It's killing it — that's our next president," said John Gagnon, 50, who had affixed an "I voted" sticker to the blue uniform shirt he wears in a betting game. "We need a change. We've got to get rid of the establishment."

### NEWS ANALYSIS

#### Around the World, Uncertainty And Fear That 'All Bets Are Off'

By PETER BAKER

JERUSALEM — Donald J. Trump's stunning election victory on Tuesday night ripped away beyond the nation's borders, sparking an international order that prevailed for decades and raising profound questions about America's place in the world.

For the first time since before World War II, Americans chose a president who promised to reverse the internationalism practiced by predecessors of both

### MAN IN THE NEWS

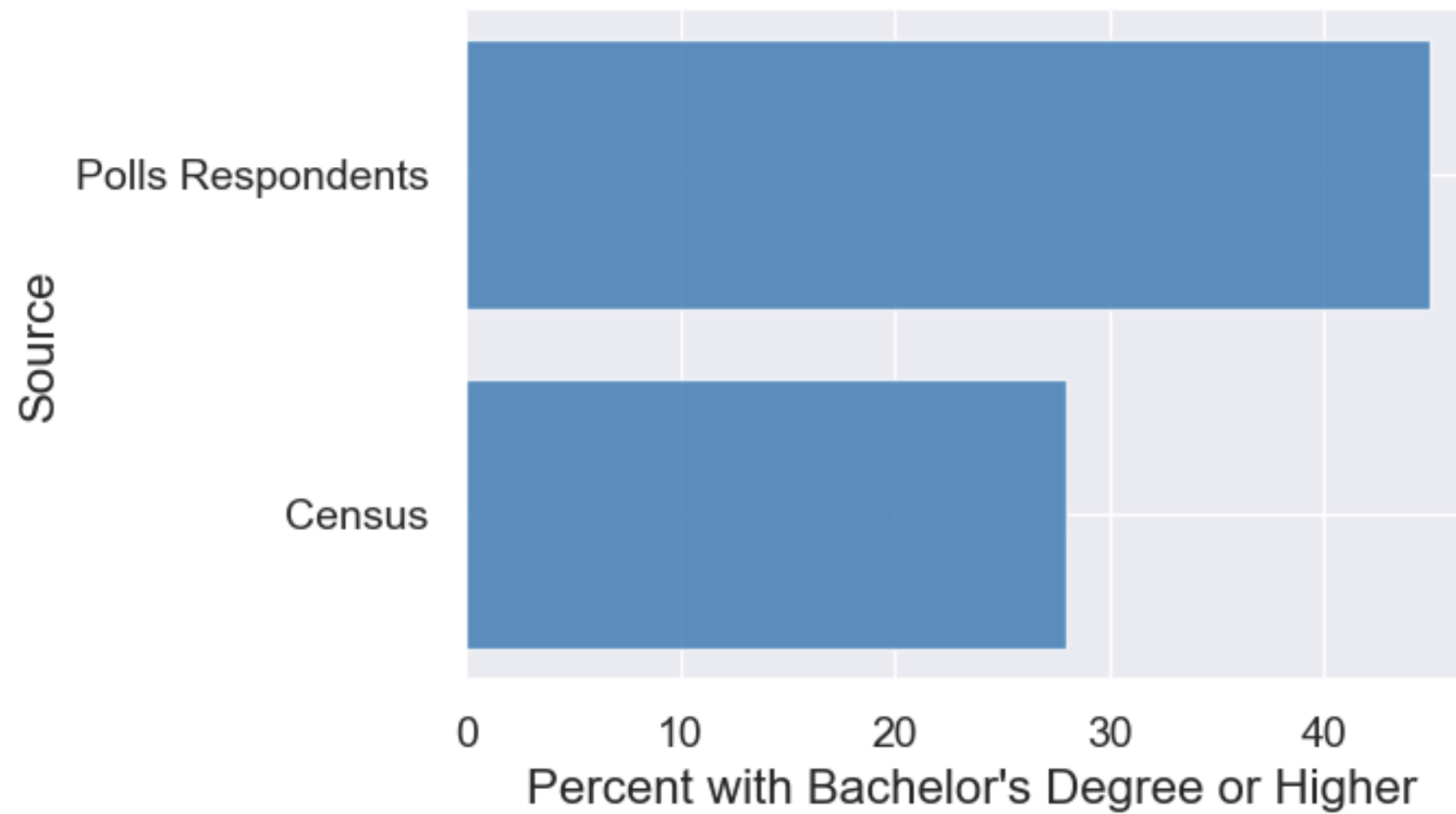
#### Clarion of White Populist Rage Who Vowed 'I Am Your Voice'

By ALEXANDER BURNS

Donald John Trump defied the skeptics who said he would never win, and the political veterans who scoffed at his slapdash campaign.

He attacked the norms of American politics, singing out groups for derision on the basis of race and religion and attacking the legitimacy of the political process. He ignored conventions of common decency, employing casual vulgarity and raising personal hostility on his political oppo-

New York Times,  
Tuesday November 8, 2016

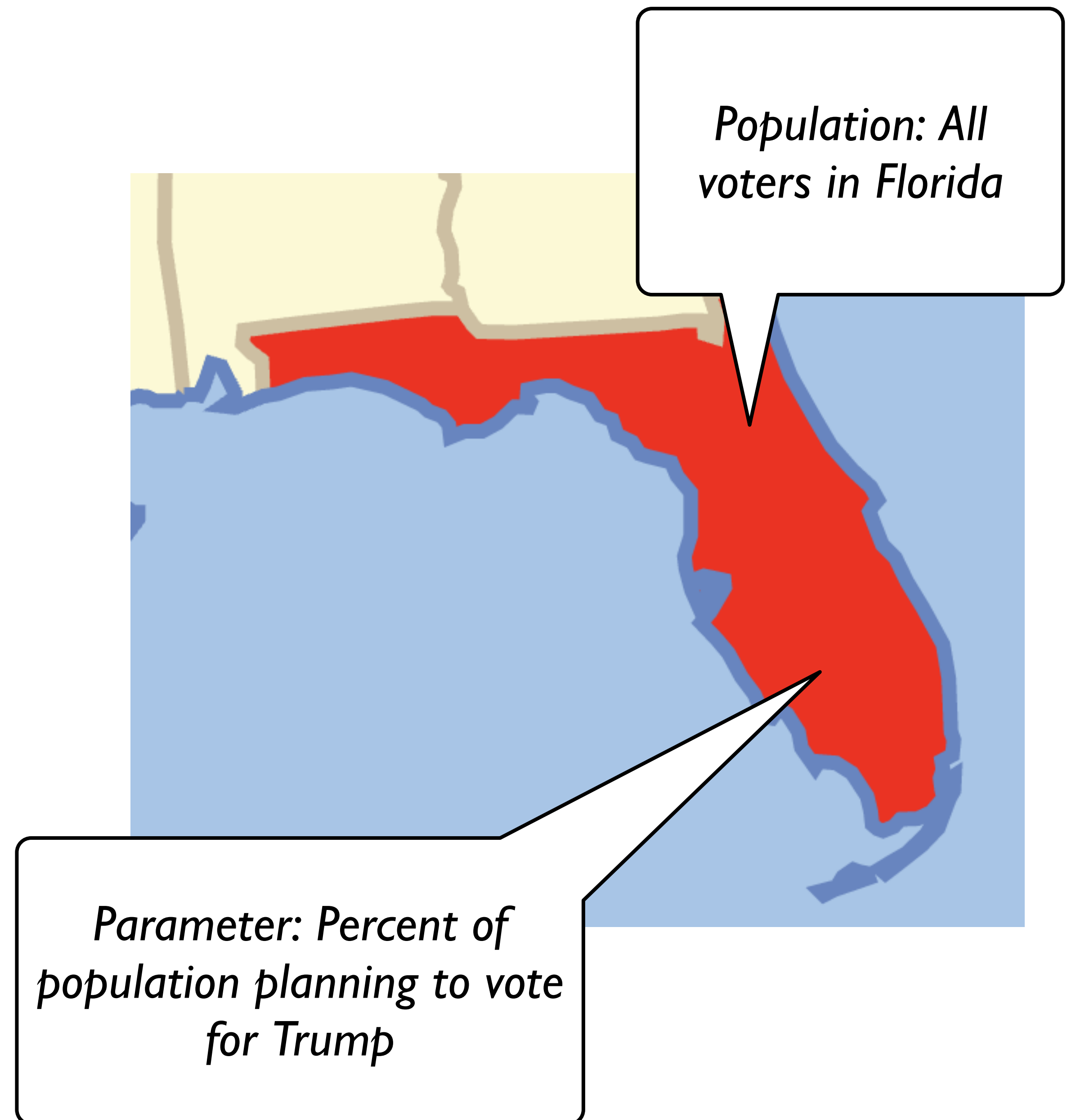


After the 2016 election, [analysis](#) showed individuals with higher education were *overrepresented* in polling samples.



# Terminology

*Parameter*: A fixed number associated with a *population*

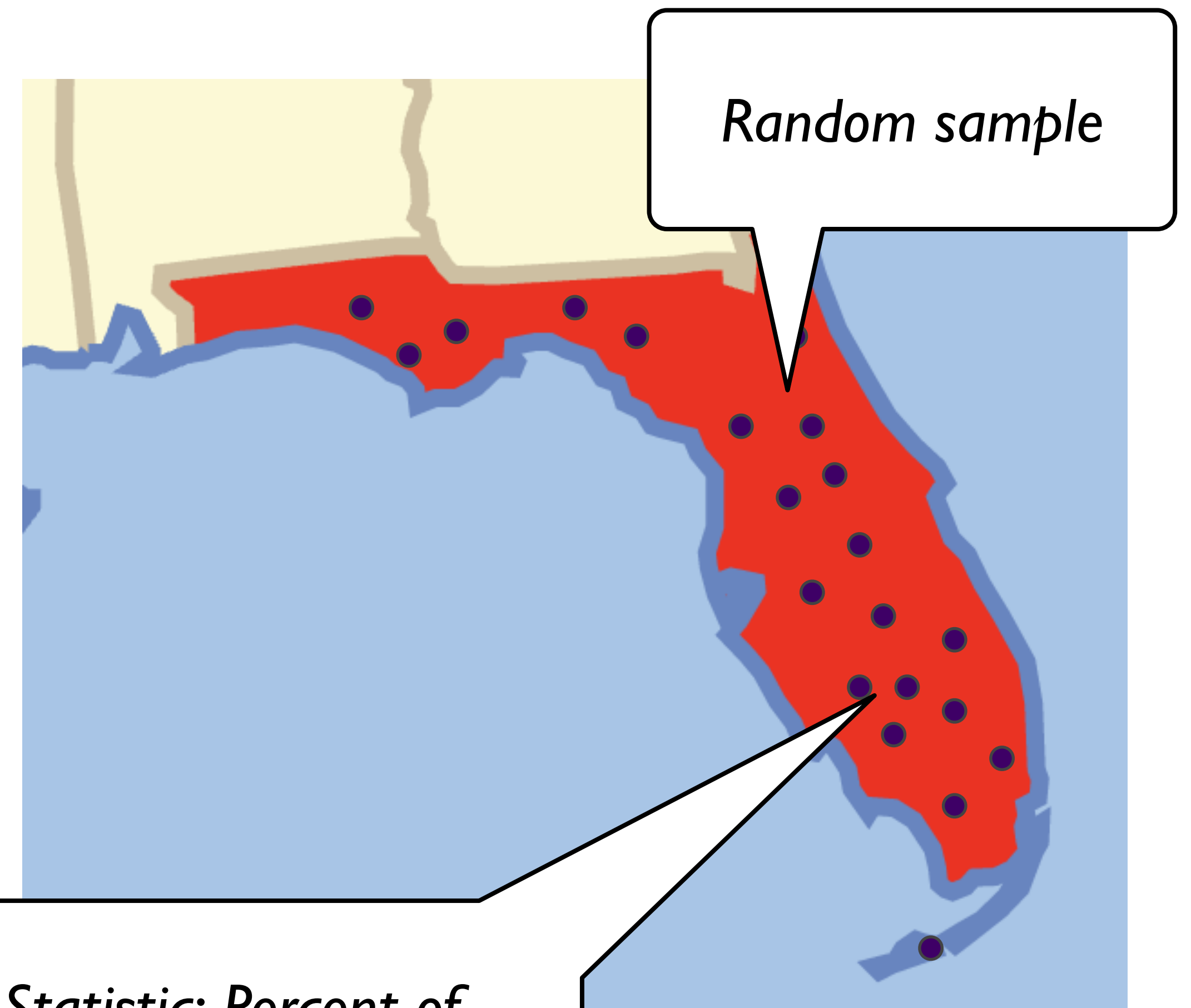


# Terminology

**Parameter:** A fixed number associated with a *population*

**Statistic:** Any number computed using the data in a sample, e.g., the mean or median.

**Statistical inference:** Estimate the value of the parameter with statistics



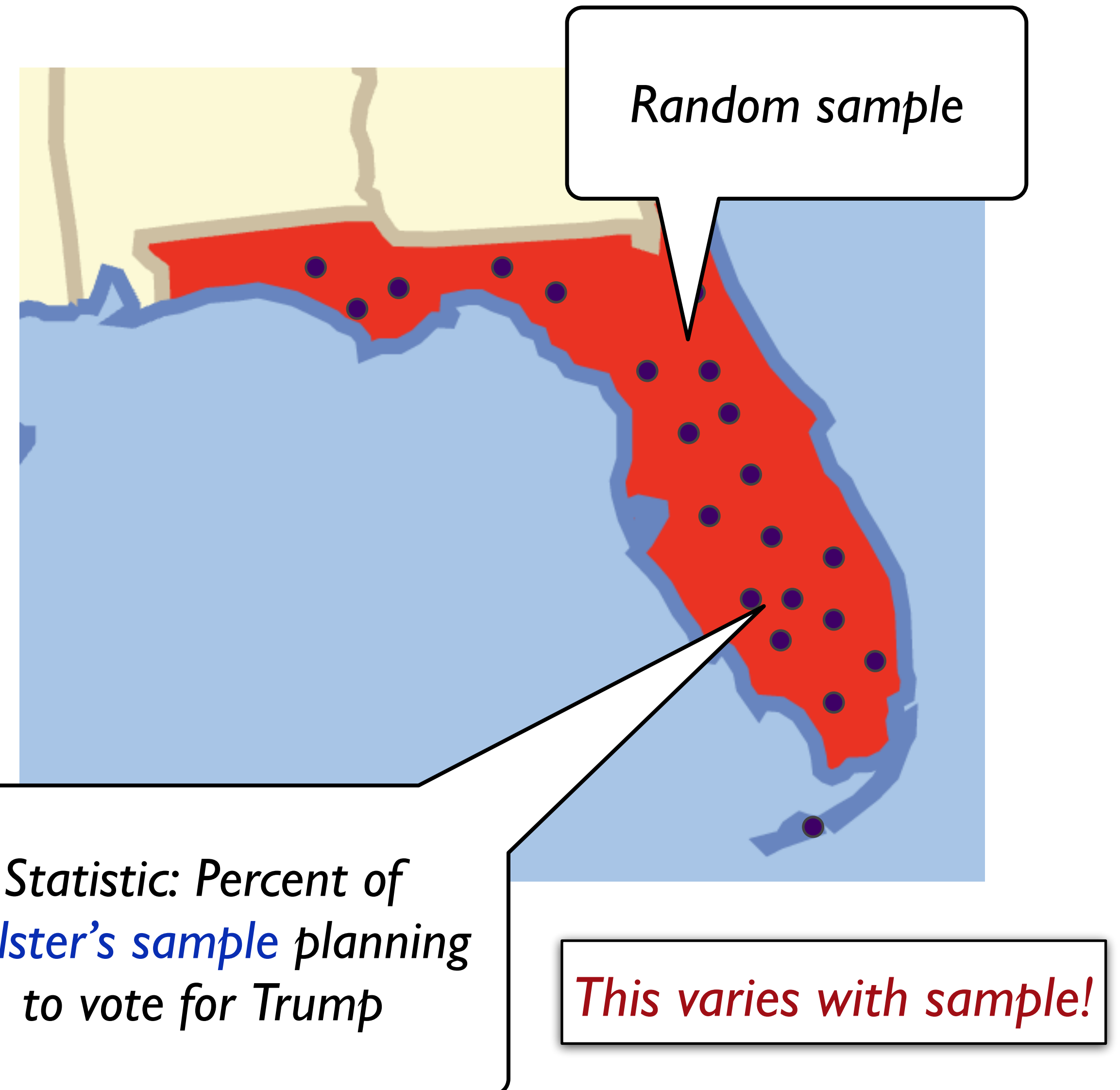
Statistic: Percent of *pollster's sample* planning to vote for Trump

# Terminology

**Parameter:** A fixed number associated with a *population*

**Statistic:** Any number computed using the data in a sample, e.g., the mean or median.

**Statistical inference:** Estimate the value of the parameter with statistics



Notebook: *Random sampling: Florida votes in 2016*

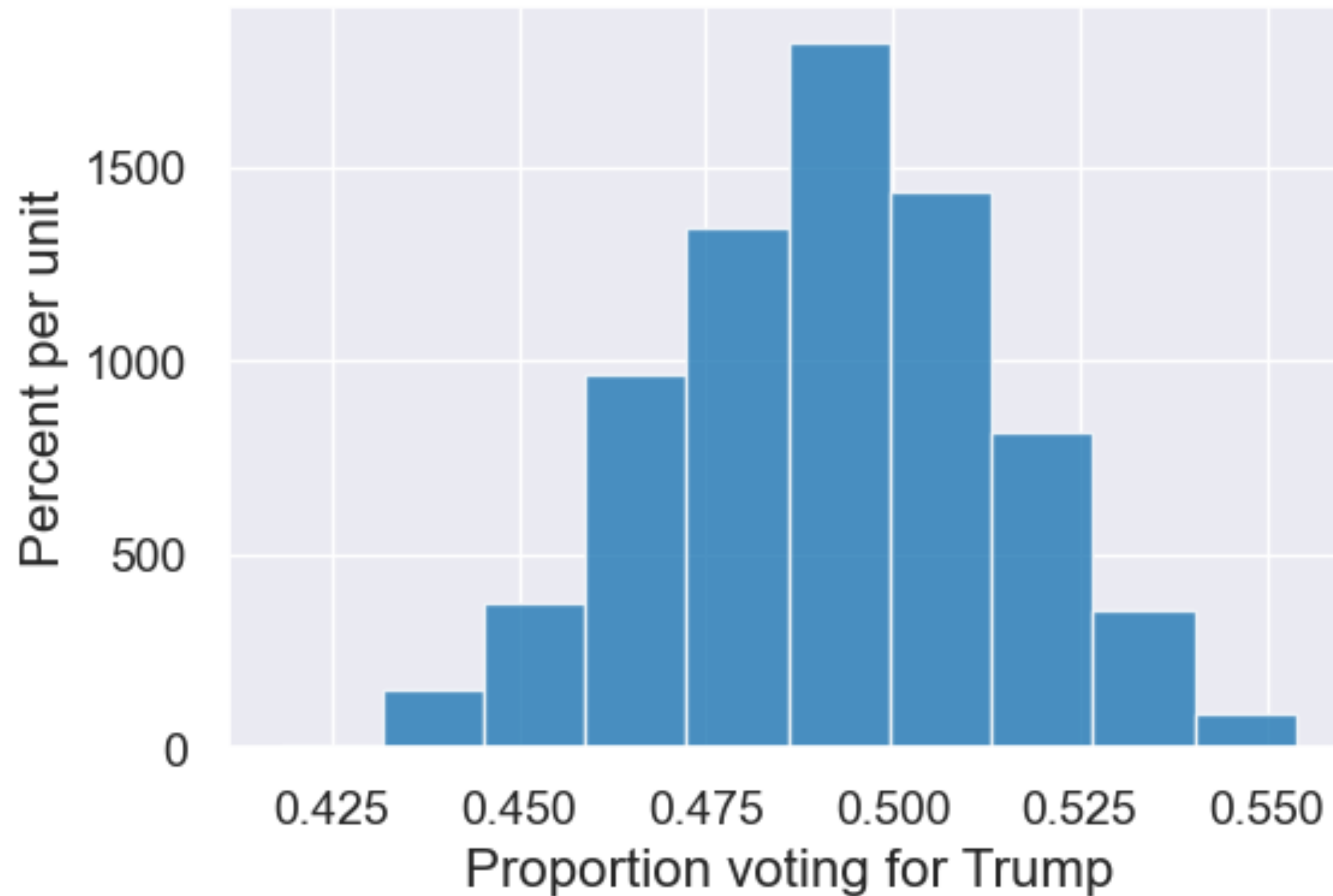
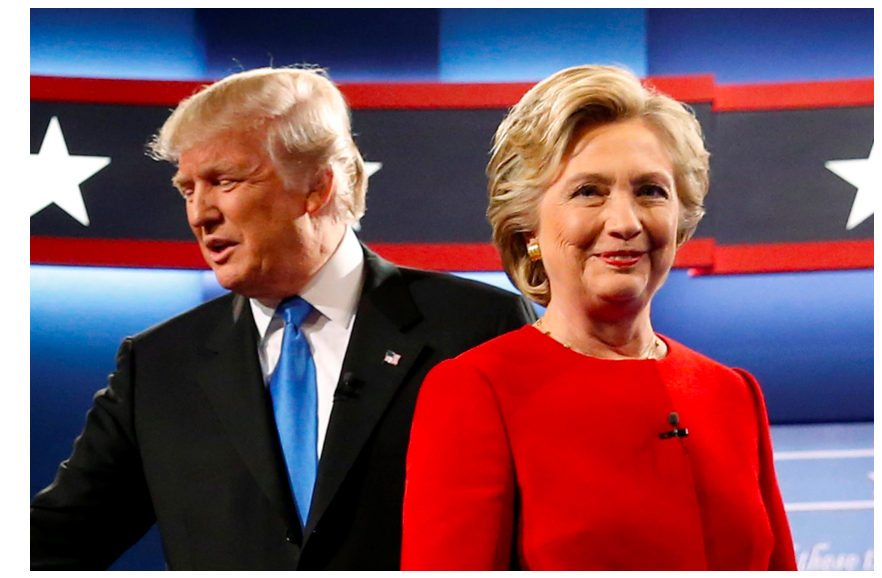
Notebook: *General sampling function*

The *sampling distribution* or probability distribution of the statistic consists of all possible values of the statistic and their corresponding probabilities.

This can be hard to calculate!

Either you need to do the math or you need to generate all possible samples and calculate the statistic based on each sample.

# Empirical distribution of a statistic



*1. Observe the statistic from repetitions of a (sampling) experiment or simulation.*

*2. Create a distribution of statistics (i.e., a histogram)*

The *empirical distribution of the statistic* is

based on simulated values of the statistic and

consists of

all the observed values of the statistic and

the proportion of times each value appeared.

The empirical distribution is a good approximation to the probability distribution of the statistic – if the number of repetitions in the simulation is large!

A fundamental consideration in using any statistic based on a *random sample* is that

the sample could have come out differently, and

therefore the statistic could have come out differently too!

538 uses polling, economic and demographic data to explore likely election outcomes.

# Trump wins 53 times out of 100

in our simulations of the 2024 presidential election.

# Harris wins 46 times out of 100.

There is a less than 1-in-100 chance of no Electoral College winner.



Trump	534
Harris	464
No winner	2
1,000 simulations	

