

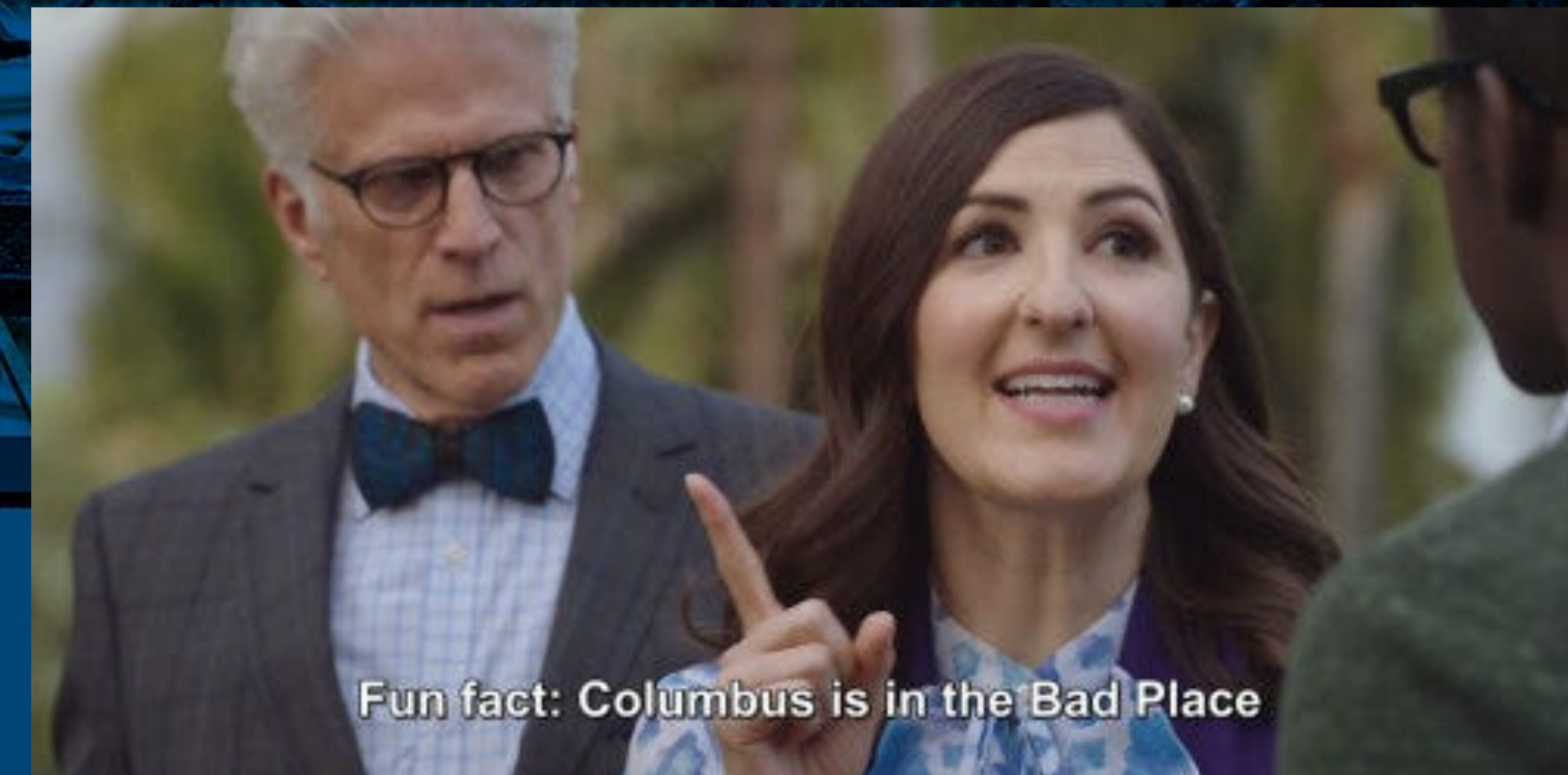
COMS BC1016

Introduction to Computational Thinking and Data Science

Lecture 10: Chance and Sampling

Sep 30, 2025

Copyright © 2026 Barnard College



October 12, 2026 1

Assignment Announcements

- Some HW dates have changed
 - Not expected to hit all topics covered in HW 4 until Wednesday
 - HW 4 due date has been pushed to Wed, Oct 28 (originally Oct 21)
 - This means HW 5 is pushed completely past Midterm
- HW dates in November will be wonky due to holidays
 - HW 6 is shorter and will be due Thanksgiving week, and HW 7 is due after Thanksgiving

Assignment Announcements

- HW Notebooks have been updated to **only** do type checks and some common errors
 - This applies to coding questions, too!
- Gradescope autograder now checks for more than what is checked in the notebook!
- Public checks passing will not necessarily correspond to 100% on the assignment
- Full results will be shown once grades are published

Other Announcements

- TAs are soliciting topics to focus on in the midterm review
- *I will be traveling Oct 21 and will not be having my regular office hours that day*
- We will start talking about Final Project after the midterm
 - This will be **groups of 2 (maybe 3)**, so I recommend starting to think who you want to work with!
 - **Your group can only consist of people from the same lecture portion**

Simulation

Experimentation

- Why do we want to run experiments?
 - Test a hypothesis
- Let's say we want to test coin flipping. What could be a hypothesis for our experiment?
 - Is the coin loaded (i.e., not fair)?

General Process for Simulations

1. Figure out what you want to simulate
 - Example: Outcomes of a coin toss
2. Write a function whose output is the outcome of a single simulation
3. Repeat the simulation for some number of iterations
 - Keep track of the results of every iteration in an array
4. Add results array to a table so you can plot the results



General Process for Simulations

1. Figure out what you want to simulate
 - Example: Outcomes of a coin toss
2. Write a function whose output is the outcome of a single simulation
3. Repeat the simulation for some number of iterations
 - Keep track of the results of every iteration in an array
4. Add results array to a table so you can plot the results



Typically what we want to know about is influenced by **chance** or **randomness**

Random Selection

To select uniformly at random from array `some_array`

- `np.random.choice(some_array)`

To select `n` number of random elements from array `some_array`

- `np.random.choice(some_array, n)`

Note: Random does not mean arbitrary.

We mean each output has some chance of happening (**probability**)

General Process for Simulations

1. Figure out what you want to simulate
 - Example: Outcomes of a coin toss
2. Write a function whose output is the outcome of a single simulation
3. Repeat the simulation for some number of iterations
 - Keep track of the results of every iteration in an array
4. Add results array to a table so you can plot the results



To keep track of our results, we will want to **append** (add to the end) elements onto our arrays

Adding Data to An Array

We can use `np.append` to add values onto the end of an array

- `np.append(array_1, value)`
 - Creates a new array with `value` appended to the end of `array_1`
- `np.append(array_1, array_2)`
 - Creates a new array with the values in `array_2` appended to the end of `array_1`

Demo: Coin Flipping

Probability

Basics of Probability

- **Lowest value:** 0
 - Chance of the event that is impossible
- **Highest value:** 1 (or 100%)
 - Chance of event that is certain
- **Complement:** The change that something *doesn't* happen
 - Computed as $(1 - \text{the chance something happens})$
 - e.g., if chance of happening is 70%, chance of not happening is 30%

Equally Likely Outcomes

Assuming all outcomes are equally likely:

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

Exercise A

- I have three cards:

Ace of Hearts, King of Diamonds, and Queen of Spades

- I shuffle them and draw two cards at random without replacement
- What's the chance that I get the Queen followed by the King?

Exercise A

What's the chance that I get the Queen followed by the King?

- Let A be event “Queen then King”

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

Exercise A

What's the chance that I get the Queen followed by the King?

- Let A be event “Queen then King”

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

↖
6

Draw 1	Draw 2
Ace	Queen
Ace	King
Queen	King
Queen	Ace
King	Ace
King	Queen

Exercise A

What's the chance that I get the **Queen** followed by the **King**?

- Let A be event “**Queen** then **King**”

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

1

6

Draw 1	Draw 2
Ace	Queen
Ace	King
Queen	King
Queen	Ace
King	Ace
King	Queen

Exercise A

What's the chance that I get the Queen followed by the King?

- Let A be event “Queen then King”

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

1
↙

6
↖

$$= \frac{1}{6}$$

Draw 1	Draw 2
Ace	Queen
Ace	King
Queen	King
Queen	Ace
King	Ace
King	Queen

Multiplication Rule

The chance that two events A and B both happen:

$$P(A \text{ and } B) = P(A) \times P(B \text{ happens given that } A \text{ happens})$$

Exercise A (another way)

What's the chance that I get the **Queen** followed by the **King**?

- Let A be event “**Queen** in the first draw”
- Let B be event “**King** in the second draw”

$$P(A \text{ and } B) = P(A) \times P(B \text{ happens given } A \text{ happens})$$

Ace

King

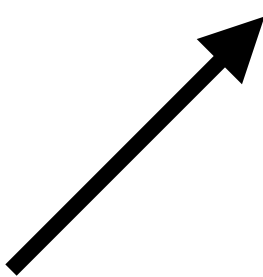
Queen

Exercise A (another way)

What's the chance that I get the **Queen** followed by the **King**?

- Let A be event “**Queen** in the first draw”
- Let B be event “**King** in the second draw”

$P(A \text{ and } B) = P(A) \times P(B \text{ happens given } A \text{ happens})$

$$\frac{1}{3}$$


Ace

King

Queen

Exercise A (another way)

What's the chance that I get the **Queen** followed by the **King**?

- Let A be event “**Queen** in the first draw”
- Let B be event “**King** in the second draw”

$$P(A \text{ and } B) = P(A) \times P(B \text{ happens given } A \text{ happens})$$

$$\frac{1}{3} \quad \nearrow \quad \frac{1}{2} \quad \nearrow$$

Ace

King

Queen

Multiplication Rule

The chance that two events A and B both happen:

$$P(A \text{ and } B) = P(A) \times P(B \text{ happens given } A \text{ happens})$$

- The answer is *less than or equal to* each of the two chances being multiplied
- The more conditions you have to satisfy, the less likely you are to satisfy them all

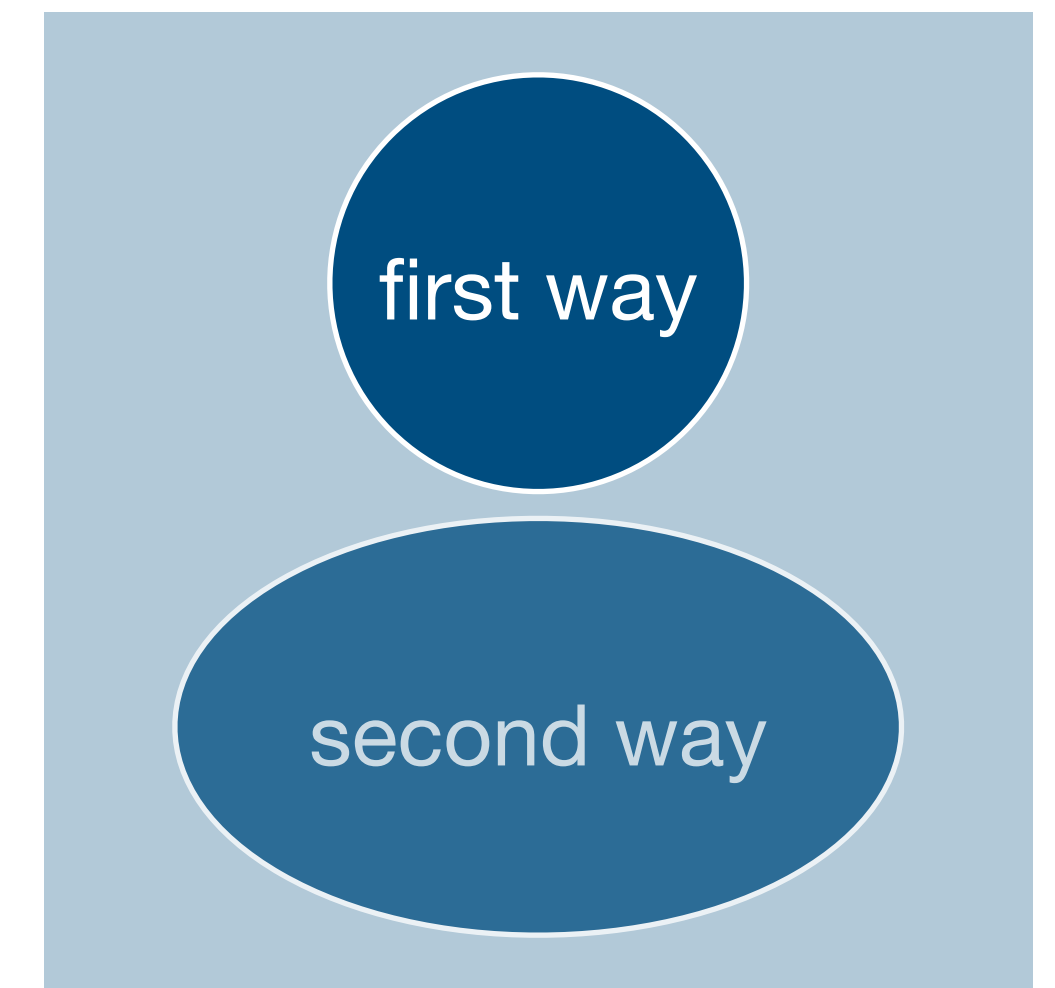
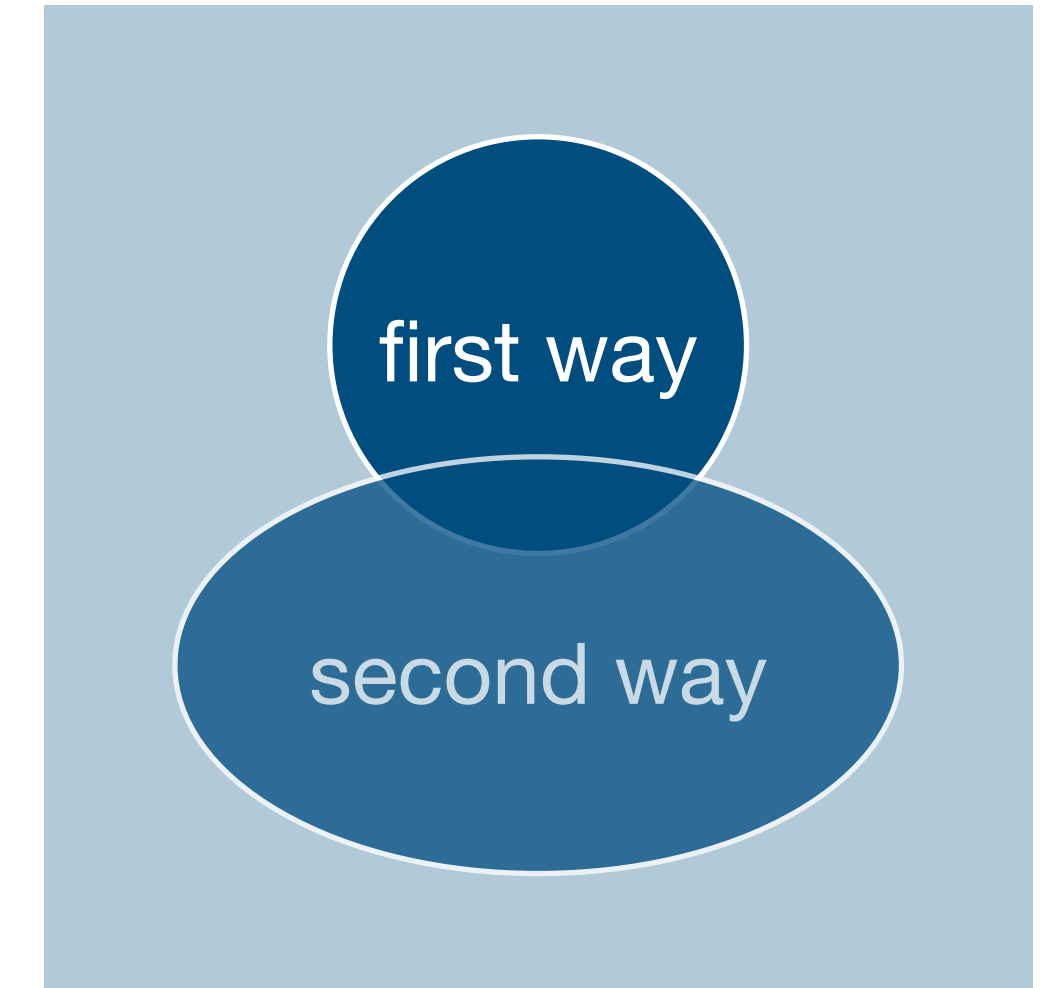
Addition Rule

- If event A can happen in exactly one of two ways, then

$$P(A) \leq P(\text{first way}) + P(\text{second way})$$

- If the two ways are independent (i.e., no overlap), then

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



In this class, we'll mostly deal with independent events

Exercise B

- I have three cards:

Ace of Hearts, King of Diamonds, and Queen of Spades

- I shuffle them and draw two cards at random without replacement
- What's the chance that one is a *Queen* and one is a *King*?

Exercise B: Equal Probabilities

What's the chance that one is the Queen and one is the King?

- Let A be drawing a Queen and King

Draw 1	Draw 2
Ace	Queen
Ace	King
Queen	King
Queen	Ace
King	Ace
King	Queen

Exercise B: Equal Probabilities

What's the chance that one is the **Queen** and one is the **King**?

- Let A be drawing a **Queen** and **King**

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

Draw 1	Draw 2
Ace	Queen
Ace	King
Queen	King
Queen	Ace
King	Ace
King	Queen

Exercise B: Equal Probabilities

What's the chance that one is the **Queen** and one is the **King**?

- Let A be drawing a **Queen** and **King**

$$P(A) = \frac{\text{number of outcomes that make } A \text{ happen}}{\text{total number of outcomes}}$$
$$= \frac{2}{6}$$

Draw 1	Draw 2
Ace	Queen
Ace	King
Queen	King
Queen	Ace
King	Ace
King	Queen

Exercise B: Equal Probabilities

What's the chance that one is the **Queen** and one is the **King**?

- Let A be drawing a **Queen** and **King**

$$\begin{aligned} P(A) &= \frac{\text{number of outcomes that make } A \text{ happen}}{\text{total number of outcomes}} \\ &= \frac{2}{6} \\ &= \frac{1}{3} \end{aligned}$$

Draw 1	Draw 2
Ace	Queen
Ace	King
Queen	King
Queen	Ace
King	Ace
King	Queen

Exercise B: Addition Rule

What's the chance that one is the Queen and one is the King?

- Let A be drawing a Queen and King

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

Draw 1	Draw 2
Ace	Queen
Ace	King
Queen	King
Queen	Ace
King	Ace
King	Queen

Exercise B: Addition Rule

What's the chance that one is the **Queen** and one is the **King**?

- Let A be drawing a **Queen** and **King**

$$\begin{aligned} P(A) &= P(\text{first way}) + P(\text{second way}) \\ &= P(\text{Queen then King}) + P(\text{King then Queen}) \end{aligned}$$

Draw 1	Draw 2
Ace	Queen
Ace	King
Queen	King
Queen	Ace
King	Ace
King	Queen

Exercise B: Addition Rule

What's the chance that one is the **Queen** and one is the **King**?

- Let A be drawing a **Queen** and **King**

$$\begin{aligned} P(A) &= P(\text{first way}) + P(\text{second way}) \\ &= P(\text{Queen then King}) + P(\text{King then Queen}) \\ &= \frac{1}{6} + \frac{1}{6} \end{aligned}$$

Draw 1	Draw 2
Ace	Queen
Ace	King
Queen	King
Queen	Ace
King	Ace
King	Queen

Exercise B: Addition Rule

What's the chance that one is the **Queen** and one is the **King**?

- Let A be drawing a **Queen** and **King**

$$\begin{aligned} P(A) &= P(\text{first way}) + P(\text{second way}) \\ &= P(\text{Queen then King}) + P(\text{King then Queen}) \\ &= \frac{1}{6} + \frac{1}{6} \\ &= \frac{1}{3} \end{aligned}$$

Draw 1	Draw 2
Ace	Queen
Ace	King
Queen	King
Queen	Ace
King	Ace
King	Queen

Exercise

A population has 100 cats including Ruby and Gertrude.
We sample 2 cats at random without replacement.

What are the following probabilities?

1. $P(\text{neither Ruby nor Gertrude are in the sample})$
2. $P(\text{both Ruby and Gertrude are in the sample})$



Pictured: Ruby and Gertrude

Exercise

A population has 100 cats including Ruby and Gertrude.
We sample 2 cats at random without replacement.

1. $P(\text{neither Ruby nor Gertrude are in the sample})$

$$P(A \text{ and } B) = P(A) \times P(B \text{ happens given } A \text{ happens})$$

Exercise

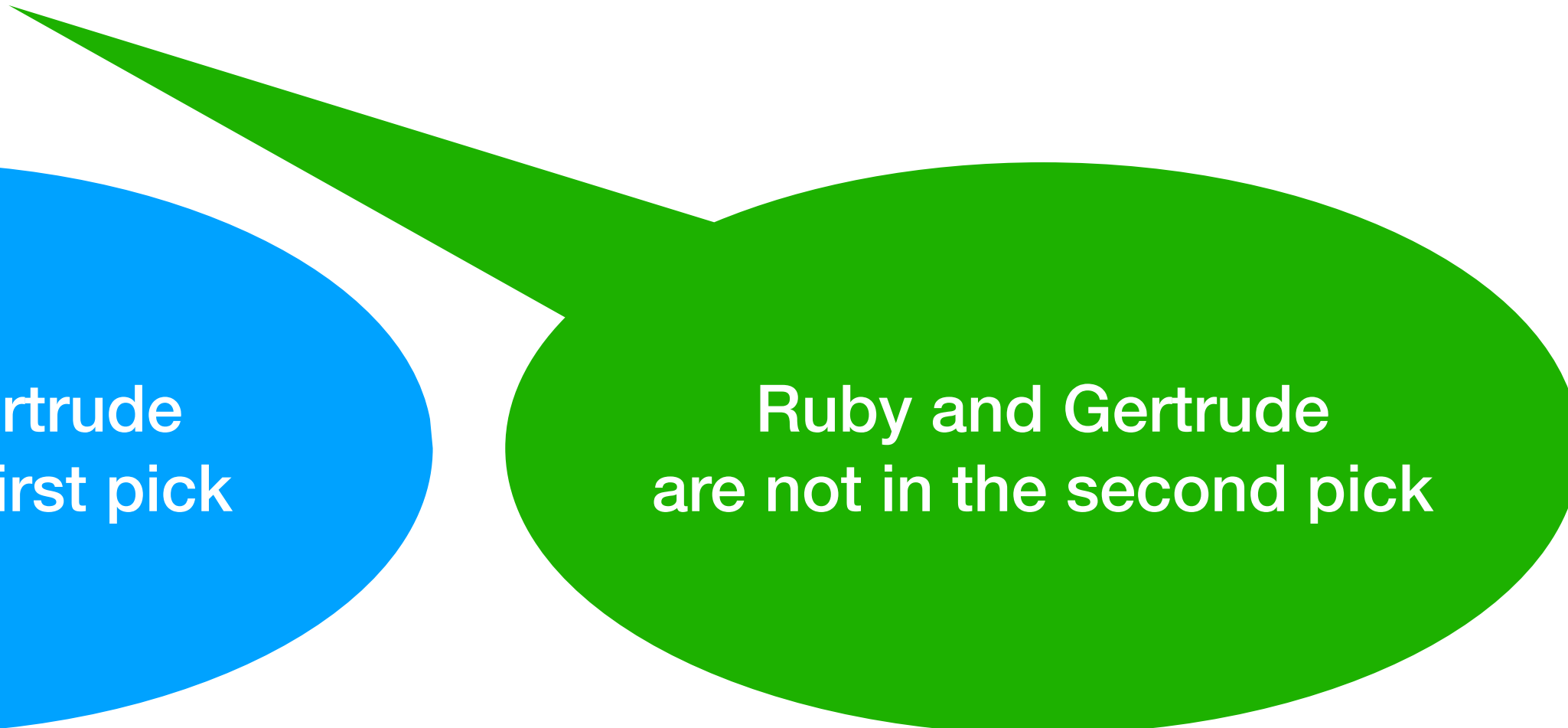
A population has 100 cats including Ruby and Gertrude.
We sample 2 cats at random without replacement.

1. $P(\text{neither Ruby nor Gertrude are in the sample})$

$$P(A \text{ and } B) = P(A) \times P(B \text{ happens given } A \text{ happens})$$



Ruby and Gertrude
are not in the first pick



Ruby and Gertrude
are not in the second pick

Exercise

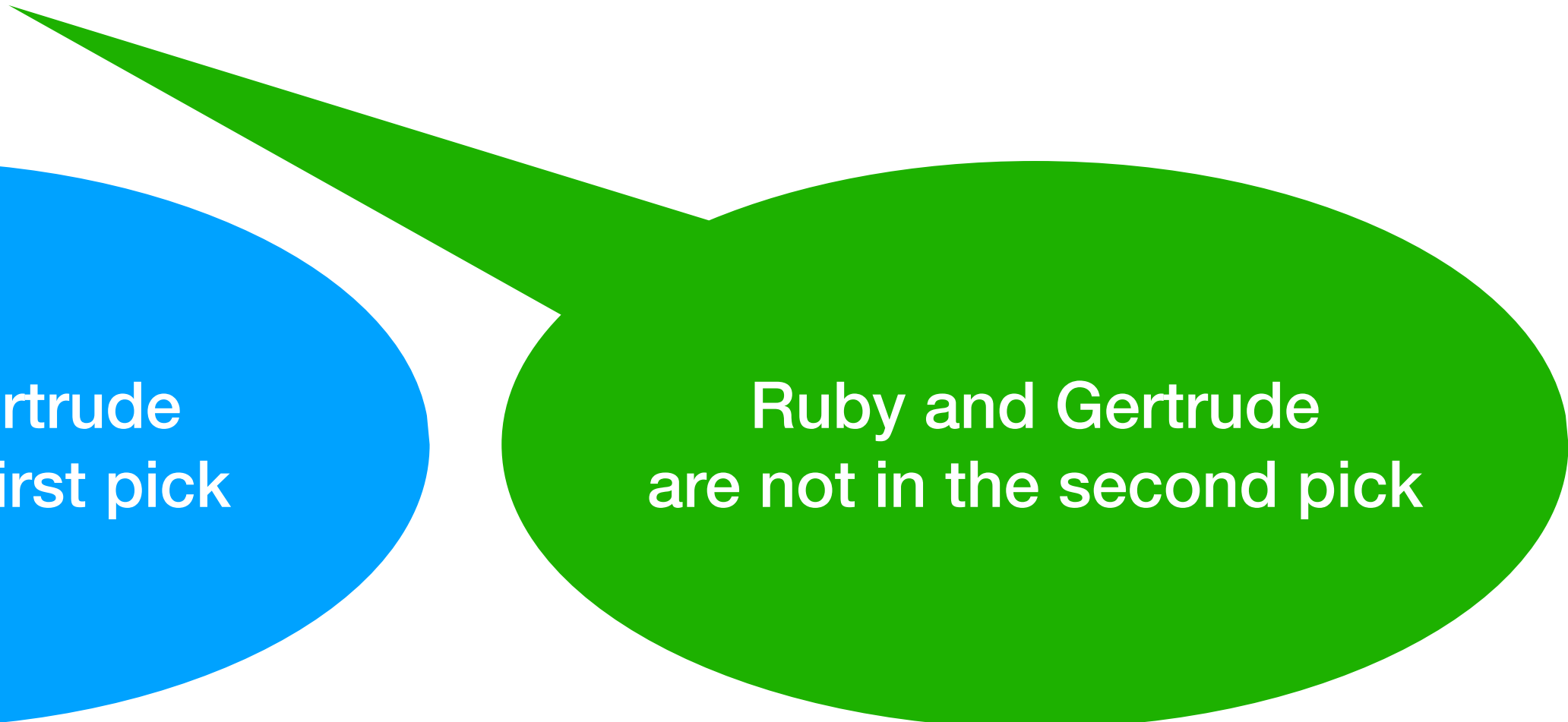
A population has 100 cats including Ruby and Gertrude.
We sample 2 cats at random without replacement.

1. $P(\text{neither Ruby nor Gertrude are in the sample})$

$$P(A \text{ and } B) = P(A) \times P(B \text{ happens given } A \text{ happens})$$



Ruby and Gertrude
are not in the first pick



Ruby and Gertrude
are not in the second pick

Exercise

A population has 100 cats including Ruby and Gertrude.
We sample 2 cats at random without replacement.

1. $P(\text{neither Ruby nor Gertrude are in the sample})$

$$\begin{aligned} P(A \text{ and } B) &= P(A) \times P(B \text{ happens given } A \text{ happens}) \\ &= \frac{98}{100} \times \frac{97}{99} \\ &= 0.96 \end{aligned}$$

Exercise

A population has 100 cats including Ruby and Gertrude.
We sample 2 cats at random without replacement.

2. $P(\text{both Ruby and Gertrude are in the sample})$

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

Exercise

A population has 100 cats including Ruby and Gertrude.
We sample 2 cats at random without replacement.

2. $P(\text{both Ruby and Gertrude are in the sample})$

$$\begin{aligned} P(A) &= P(\text{first way}) + P(\text{second way}) \\ &= P(\text{Ruby then Gertrude}) + P(\text{Gertrude then Ruby}) \end{aligned}$$

Exercise

A population has 100 cats including Ruby and Gertrude.
We sample 2 cats at random without replacement.

2. $P(\text{both Ruby and Gertrude are in the sample})$

$$\begin{aligned} P(A) &= P(\text{first way}) + P(\text{second way}) \\ &= P(\text{Ruby then Gertrude}) + P(\text{Gertrude then Ruby}) \\ &= \frac{1}{100} \times \frac{1}{99} + \frac{1}{100} \times \frac{1}{99} \\ &= 0.0002 \end{aligned}$$

Sampling

Class Survey

<https://forms.gle/WTAs8KC5TunVMkmA6>



Sample

A **sample** is a **subset of a population** you choose to utilize in your analysis

- Picking samples is a fundamental part of Data Science
 - Did you sample enough / collect enough data?
 - Is the data representative?

Deterministic vs Random Samples

Deterministic Sample: Sampling scheme doesn't involve chance, results are always the same

- Examples:
 - First 100 students when listed in alphabetical order
 - `cat_tbl.where('Coloring', 'tuxedo')`

Name	Age	Weight	Coloring	Sex	Owner
Ruby	14	8	tuxedo	F	Alice
Gertrude	15	12	tuxedo	F	Alice
Hamby	8	16	tabby	M	Bob
Fig	3	7	tabby	F	Bob
Corina	6	10	tortie	F	Carol
Frito	2	8.5	tabby	M	Carol

Deterministic vs Random Samples

Random Sample: Each element has a probability of being chosen

- Selection probabilities for each element are known *before the sample is drawn*
- Not all individuals or groups have to have equal chance of being selected
 - Example: drawing a face card vs a numbered card
- Example: `np.random.choice(np.arange(10))`



Randomly Selecting from Arrays

To select uniformly at random from array `some_array`

- `np.random.choice(some_array)`

To select `n` number of random elements from array `some_array`

- `np.random.choice(some_array, n)`

Randomly Sampling Tables

Returns a table with `n` rows sampled *with replacement* from Table `tbl`

- `tbl.sample(n)`

Returns a new table with `n` rows sampled *without replacement* from `tbl`

- `tbl.sample(n, with_replacement=False)`

Demo: Sampling

Convenience Sampling

Random sampling requires knowing the probability of selection *ahead of time*

- Not fully controlling selection doesn't necessarily make it a random sample

If you can't figure out ahead of time

- what's the population
- what's the chance of selection for each group in the population

then it is a **sample of convenience** and not a random sample!

Motivating Example

Suppose a pharmaceutical company is conducting a clinical trial for a new diabetes drug.

They want to draw a random sample of patients from the general population to test how effective the drug is.

Goal of random sampling is to make sure results can be generalized to the whole population

- To do this accurately, we need to know the selection probability of every subgroup (e.g., age, gender, race, socioeconomic status, ...) in the population

What can go wrong?

Suppose the company only recruits participants from wealthier urban neighborhoods.

What groups may be over/under-represented?

Overall sample is *biased*

- Selection probability was effectively zero for certain groups
- Claims about the drug apply to the people sampled but not necessarily the population as a whole

Next Class

- Today
 - Simulations
 - Probability
 - Sampling
- Wednesday
 - Distributions
 - Models