

COMS BC1016

Introduction to Computational Thinking and Data Science

## Lecture 10: Probability and Sampling

# Upcoming Schedule

Today

Date	Topic	Lab	Assignment
10/1	9 - Conditionals and Iteration	Lab 4 - Functions and Visualizations (Due 10/3) <a href="#">Courseworks</a>	HW2 Due
10/6	10 - Probability and Sampling		HW4 - Probability, Simulation, Estimation (Due 10/15) <a href="#">Courseworks</a>
10/8	11 - Models and Empirical Simulations	Lab 5 - Simulations (Due 10/10) <a href="#">Courseworks</a>	HW3 Due
10/13	Programming/Python Review		
10/15	Midterm Review	No Lab	HW4 Due
10/20	<b>Midterm Exam</b>		
10/22	Special Topics - Bias in AI	No Lab	

# Lecture Outline

- Control Statements
  - For loops
- Randomness
- Probabilities
- Sampling

# **Control Statements**

# Control Statements

**Control Statements** modify *if* and/or *how many times* a block of code is executed in a program

# Control Statements

- Two major types are **if** and **for**
  - **if** statements specify code that should be run conditioned on something being true
    - They can also specify if alternative code should be run otherwise
  - **for** loops allow executing code over each element in some sequence of items

# if statements

- Conditionals begin with an `if` followed by a boolean statement
  - Runs code based on whether a boolean statement evaluates to `True`
- Conditionals can include a combination of `if`, `elif`, and `else` clauses
  - Maximum of one `if` and one `else`

# if statements

```
if statement_1:  
    first_code_block  
  
elif statement_2:  
    second_code_block  
  
elif statement_3:  
    third_code_block  
  
else:  
    fourth_code_block
```

# if statements

```
if statement_1:
```

```
    first_code_block
```

Runs if statement\_1 == True

```
elif statement_2:
```

```
    second_code_block
```

Runs if statement\_1 != True  
AND statement\_2 == True

```
elif statement_3:
```

```
    third_code_block
```

statement\_1 != True  
AND statement\_2 != True  
AND statement\_3 == True

```
else:
```

```
    fourth_code_block
```

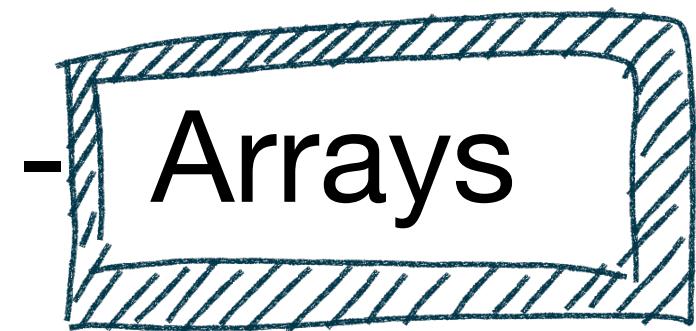
nothing above == True

# Iteration

- **Iteration** means to repeat a process or steps
  - For example, coming up with a design, prototyping, testing, and then repeating these steps based on the outcome
- In programming we use this term to refer to executing code repeatedly over every element in a list/array/sequence/collection/...
  - The object being iterated over is referred to as an **iterable**

# Iterables

- Formally, an iterable is any Python object capable of returning its members one at a time
- Iterables we've seen in this class include:



- Lists
- String

We'll mostly focus  
on arrays

```
make_array('a', 'b', 'c', 'd')
```

```
array(['a', 'b', 'c', 'd'],  
      dtype='<U1')
```

```
['a', 'b', 'c', 'd']
```

```
['a', 'b', 'c', 'd']
```

```
'abcd'
```

```
'abcd'
```

# for Statements

- Executing a **for** runs code with each element in an iterable

A diagram illustrating a `for` loop. It shows the `for` keyword, followed by a variable name `item` in a pink box, followed by `in`, and then an array of values `some_array` in a blue box. Handwritten annotations with arrows point from the text "variable name" to the word `item`, and from the text "array of values" to the box containing `some_array`.

```
for item in some_array:
```

```
    print(item)
```

A diagram of a `for` loop structure. The loop definition `for item in some_array:` is at the top. Below it is a yellow box containing the code `print(item)`. A handwritten annotation with a curved arrow points from the text "code to evaluate in each iteration of the loop" to the `print(item)` line.

# for Example

```
for i in np.arange(4):  
    print('iteration', i)
```

# for Example

```
for i in np.arange(4):  
    print('iteration', i)
```

```
np.arange(4)
```

```
array([0, 1, 2, 3])
```

# for Example

```
for i in np.arange(4):  
    print('iteration', i)
```

```
iteration 0
```

```
np.arange(4)
```

```
array([0, 1, 2, 3])
```



i=0

# for Example

```
for i in np.arange(4):  
    print('iteration', i)
```

```
iteration 0  
iteration 1
```

```
np.arange(4)
```

```
array([0, 1, 2, 3])
```



i=1

# for Example

```
for i in np.arange(4):  
    print('iteration', i)
```

```
iteration 0  
iteration 1  
iteration 2
```

```
np.arange(4)
```

```
array([0, 1, 2, 3])
```



i=2

# for Example

```
for i in np.arange(4):  
    print('iteration', i)
```

```
iteration 0  
iteration 1  
iteration 2  
iteration 3
```

```
np.arange(4)
```

```
array([0, 1, 2, 3])
```



i=3

# for Example

```
total = 0
for i in np.arange(4):
    total = total + i
print(total)
```

```
np.arange(4)
array([0, 1, 2, 3])
```

# for Example

```
total = 0
for i in np.arange(4):
    total = total + i
print(total)
```

0

```
np.arange(4)
array([0, 1, 2, 3])
```



i=0

# for Example

```
total = 0
for i in np.arange(4):
    total = total + i
print(total)
```

0  
1

```
np.arange(4)
array([0, 1, 2, 3])
```



i=1

# for Example

```
total = 0
for i in np.arange(4):
    total = total + i
print(total)
```

0  
1  
3

```
np.arange(4)
array([0, 1, 2, 3])
```

↑  
i=2

# for Example

```
total = 0
for i in np.arange(4):
    total = total + i
print(total)
```

0  
1  
3  
6

```
np.arange(4)
array([0, 1, 2, 3])
```

↑  
i=3

# Simulation

# General Process for Simulations

1. Figure out what you want to simulate
  - Example: Outcomes of a coin toss



# 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



# 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



# 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



# Random Selection

```
import numpy as np
```

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*)

# Appending Arrays

```
import numpy as np
```

Return a copy of array\_1 where value is added onto the end

```
np.append(array_1, value)
```

Returns an array with elements of array\_1 followed by elements of array\_2

```
np.append(array_1, array_2)
```

# Probability

# Probability

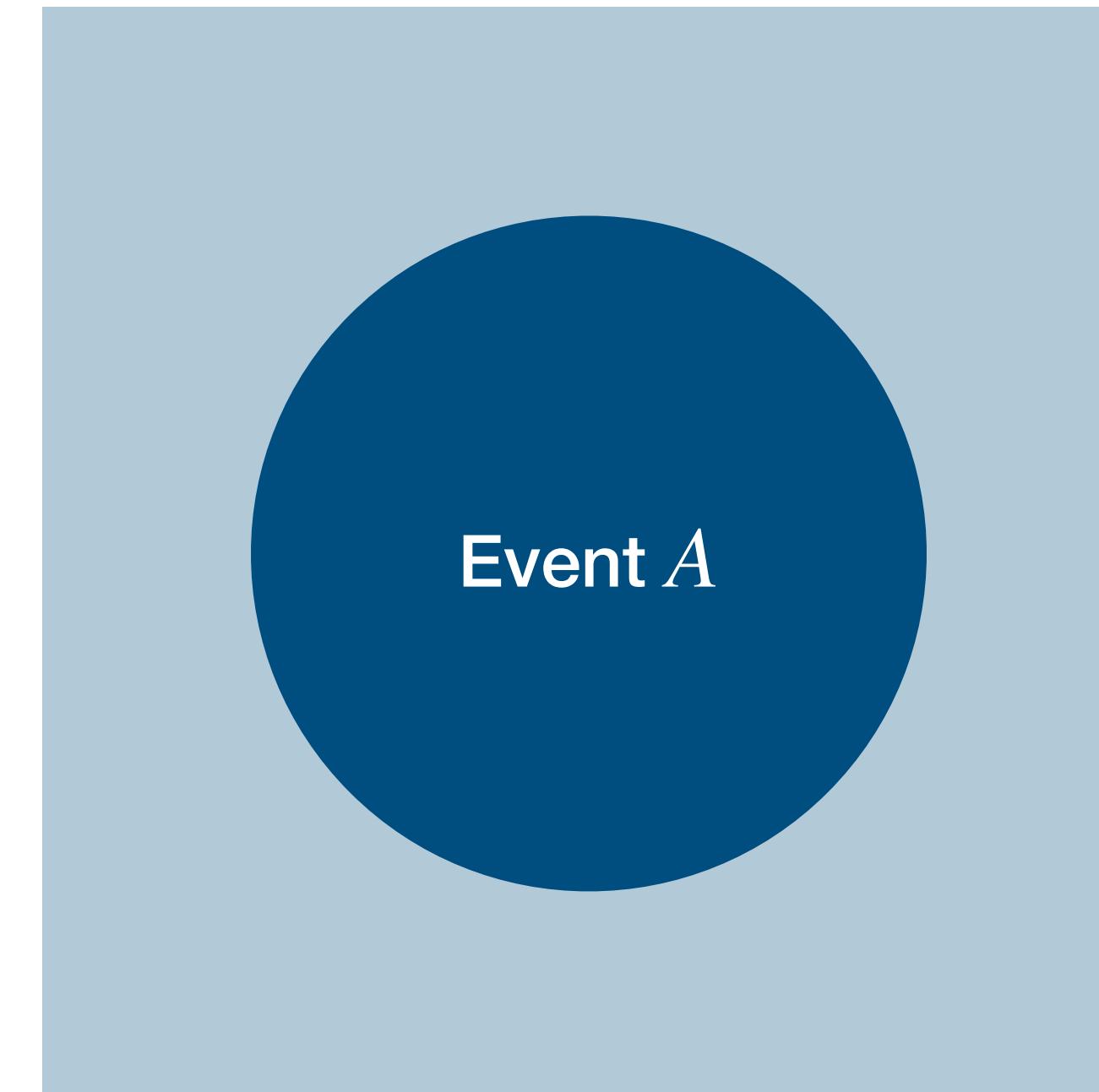
$P(A)$  = Probability of event  $A$  happening

- $P(A) = 0$ 
  - 0% chance of event  $A$  happening
  - Event  $A$  is impossible
- $P(A) = 1$ 
  - 100% chance of event  $A$  happening
  - Event  $A$  is certain

# Complements

If an event has a chance of happening  $N$ , then the chance it *doesn't* happen is  $1-N$

- 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

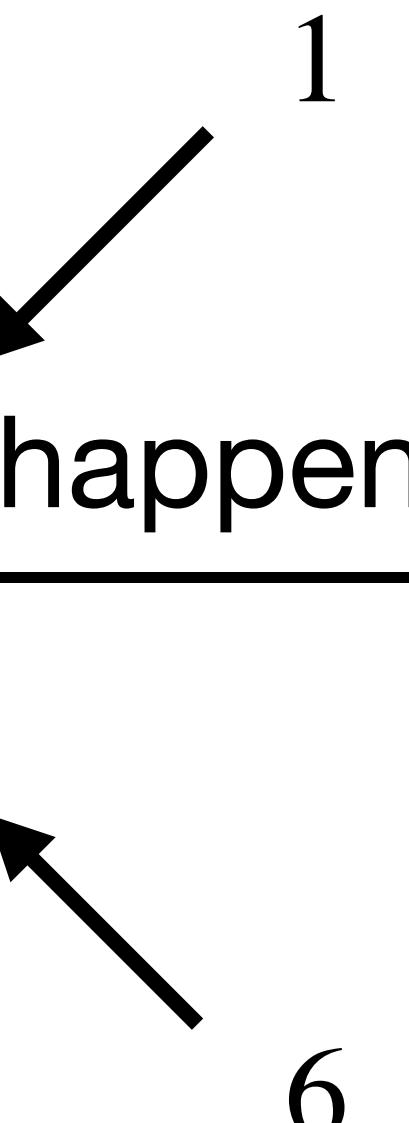
6

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”

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


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:

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

# 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”

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

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

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

$$\frac{2}{6}$$

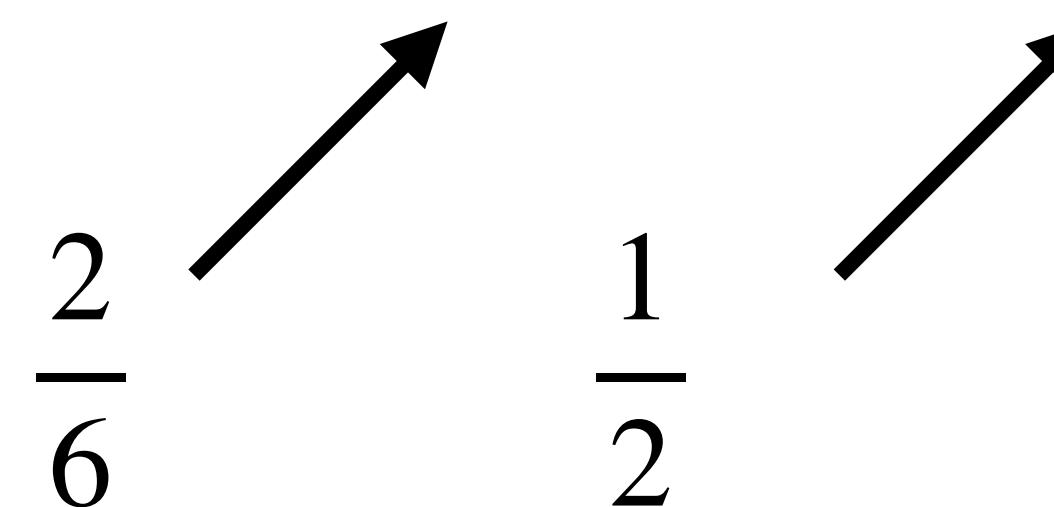
Draw 1	Draw 2
Ace	Queen
Ace	King
Queen	King
Queen	Ace
King	Ace
King	Queen

# 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”

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

$$\frac{2}{6} \quad \frac{1}{2}$$


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:

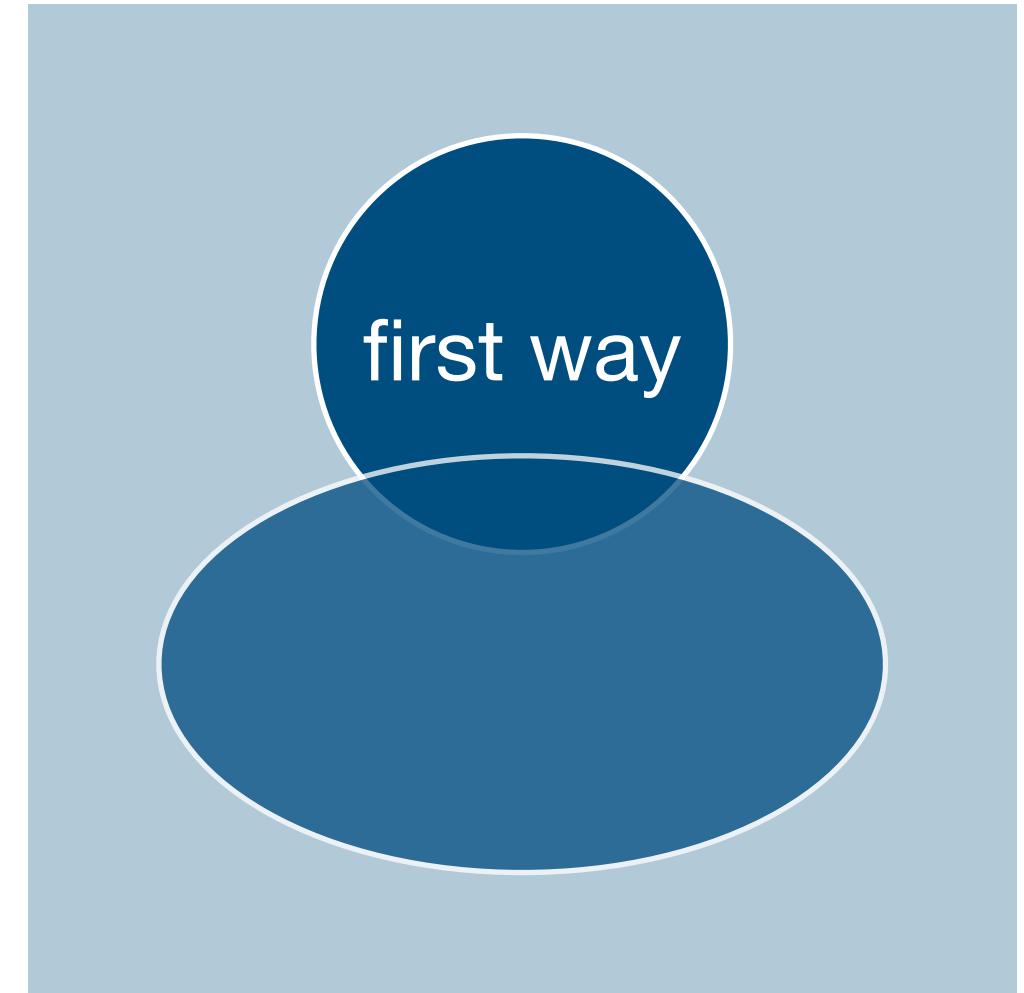
$$P(A \text{ and } B) = P(A) \times P(B \text{ happens given } A \text{ happens})$$

- 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

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



# Addition Rule

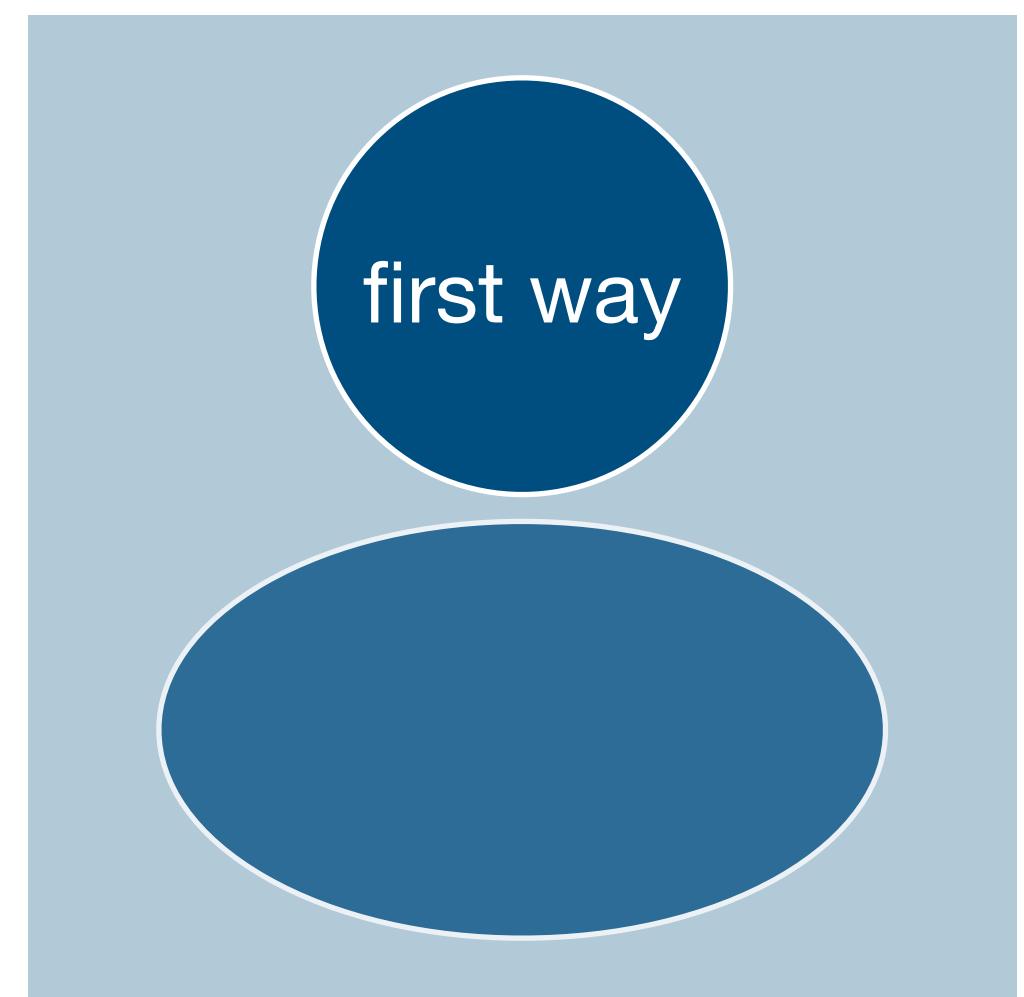
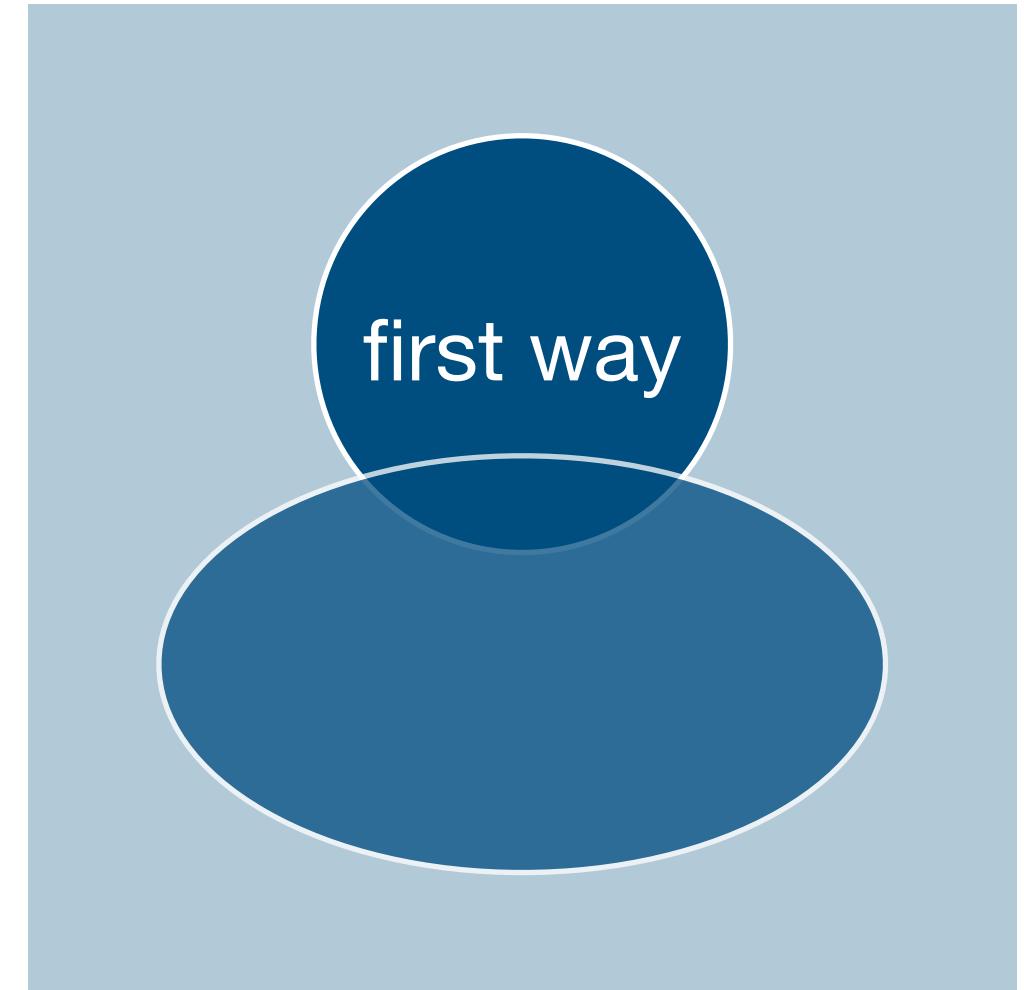
- If event A can happen in exactly one of two ways, then

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

- 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

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

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

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

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

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

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

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

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

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

First pick:  
100 cats  
98 are not Ruby or Gertrude  
 $\hookrightarrow P(A) = \frac{98}{100}$

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

$\nearrow$   
A: Gertrude and Ruby are not chosen  
in the first pick

# Exercise

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

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

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

$\hookrightarrow$  A: Gertrude and Ruby are not chosen  
in the first pick

$\hookrightarrow$  B: Gertrude and Ruby are not chosen  
in the second pick

First pick:

100 cats

98 are not Ruby or Gertrude

$$\hookrightarrow P(A) = \frac{98}{100}$$

Second pick:

99 cats after A (removed 1)

97 are not Ruby or Gertrude

$$\hookrightarrow P(B \text{ given } A) = \frac{97}{99}$$

# Exercise

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

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

First pick:  
100 cats  
98 are not Ruby or Gertrude  
 $\hookrightarrow P(A) = \frac{98}{100}$

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

$$= \frac{98}{100} \times \frac{97}{99}$$
$$= 0.96$$

Second pick:  
99 cats after A (removed 1)  
97 are not Ruby or Gertrude  
 $\hookrightarrow P(B \text{ given } A) = \frac{97}{99}$

# Exercise

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

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

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

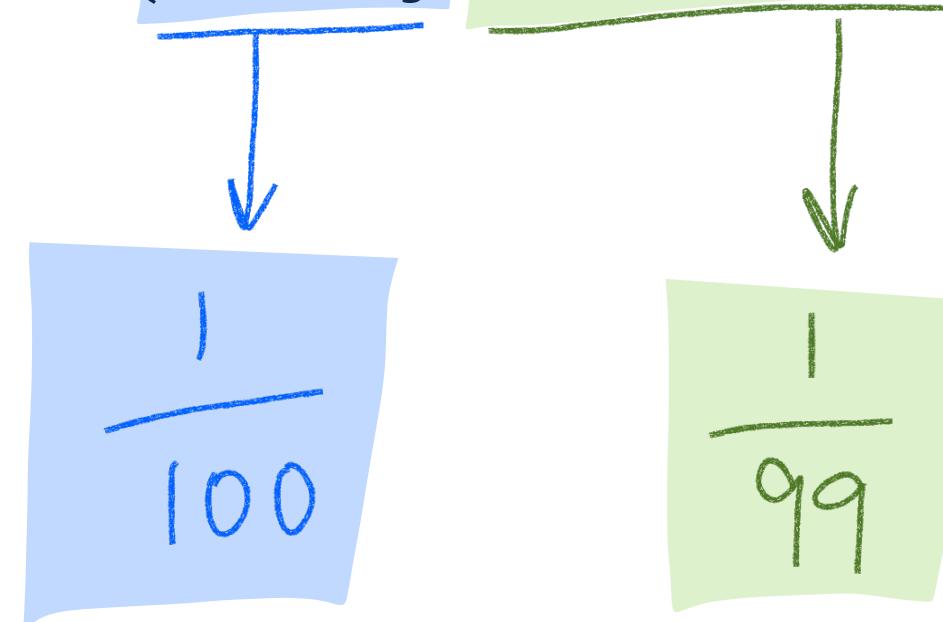
# Exercise

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

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

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

$$= P(\text{Ruby then Gertrude}) + P(\text{Gertrude then Ruby})$$

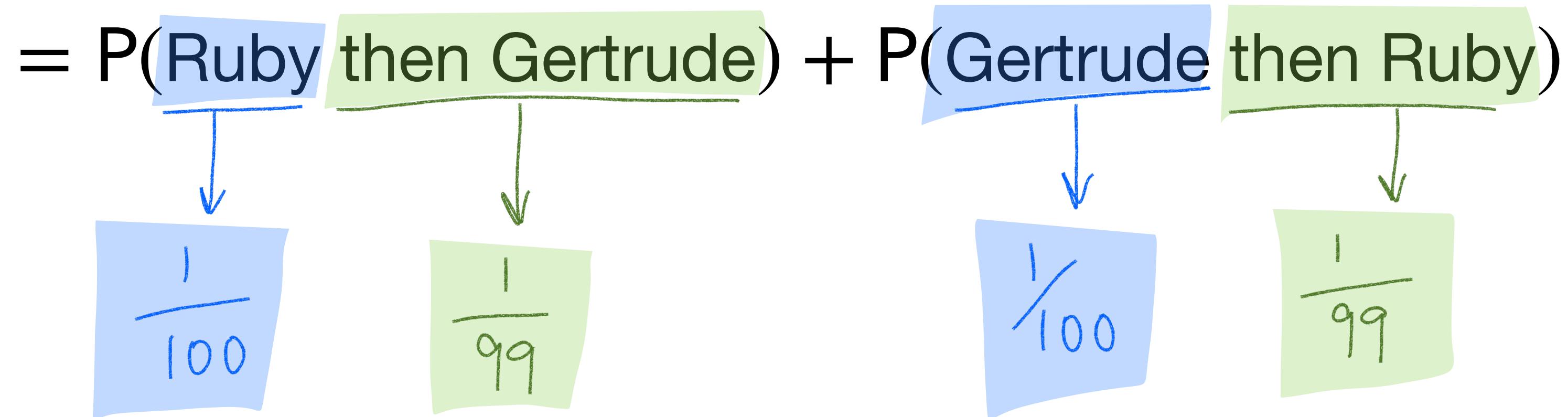


# Exercise

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

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

# **Sampling**

# Sample

- A subset of your 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
  - Example: `cat_tbl.where('Coloring', 'tuxedo')`
- Random Sample:
  - Elements are chosen probabilistically
    - 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: `np.random.choice(np.arange(10))`

# Convenience Sampling

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

- Not fully controlling what is picked (e.g., “choose the first 10 people to walk by”) 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!

# Next Time

Wed

Date	Topic	Lab	Assignment
10/1	9 - Conditionals and Iteration	Lab 4 - Functions and Visualizations (Due 10/3) <a href="#">Courseworks</a>	HW2 Due
10/6	10 - Probability and Sampling		HW4 - Probability, Simulation, Estimation (Due 10/15) <a href="#">Courseworks</a>
10/8	11 - Models and Empirical Simulations	Lab 5 - Simulations (Due 10/10) <a href="#">Courseworks</a>	HW3 Due
10/13	Programming/Python Review		
10/15	Midterm Review	No Lab	HW4 Due
10/20	<b>Midterm Exam</b>		
10/22	Special Topics - Bias in AI	No Lab	