University of Southampton

# BAYESIAN CREATURES

# L2: SETS FROM DATA

# LESSON OUTCOMES

LO 1: Give definitions for what is meant by a **set union**, and **set intersection**.

LO 2: Construct probabilities from sets

IDEA

Suppose we collected data from a lot of animals

| ID | Name | Eating Habits | Primary Domain | Endangered Status |
|---|---|---|---|---|
| a | African Forest Elephant | Herbivore | Land | Critical |
| b | Bearded Vulture | Carnivore | Air | Critical |
| c | Condor | Carnivore | Air | Critical |
| d | Dugong | Omnivore | Water | Vulnerable |
| e | Emperor Penguin | Carnivore | Land | Critical |

**IDEA**

This is our superset, the animals. We'll put their IDs in a set called **A**.

| ID | Name | Eating Habits | Primary Domain | Endangered Status |
|---|---|---|---|---|
| a | African Forest Elephant | Herbivore | Land | Critical |
| b | Bearded Vulture | Carnivore | Air | Critical |
| c | Condor | Carnivore | Air | Critical |
| d | Dugong | Omnivore | Water | Vulnerable |
| e | Emperor Penguin | Carnivore | Land | Critical |

## MAIN SET

$$A = \{a, b, c, d, e\}$$

**IDEA**

The animals eating habits give us three subsets. **H**, **C**, and **O**.

| ID | Name | Eating Habits | Primary Domain | Endangered Status |
|---|---|---|---|---|
| a | African Forest Elephant | Herbivore | Land | Critical |
| b | Bearded Vulture | Carnivore | Air | Critical |
| c | Condor | Carnivore | Air | Critical |
| d | Dugong | Omnivore | Water | Vulnerable |
| e | Emperor Penguin | Carnivore | Land | Critical |

# EATING SUBSETS

$H = \{a\}$

$C = \{b, c, e\}$

$O = \{d\}$

**IDEA**

The animals’ primary domains give us three more subsets. **L**, **W**, and **Air**.

| ID | Name | Eating Habits | Primary Domain | Endangered Status |
|---|---|---|---|---|
| a | African Forest Elephant | Herbivore | Land | Critical |
| b | Bearded Vulture | Carnivore | Air | Critical |
| c | Condor | Carnivore | Air | Critical |
| d | Dugong | Omnivore | Water | Vulnerable |
| e | Emperor Penguin | Carnivore | Land | Critical |

# DOMAIN SUBSETS

$$L = \{a, e\}$$

$$W = \{d\}$$

$$Air = \{b, c\}$$

## QUESTION 1

Which subsets do the animals’ endangered status give us? Write them out.

| ID | Endangered Status |
|---|---|
| a | Critical |
| b | Critical |
| c | Critical |
| d | Vulnerable |
| e | Critical |

## ANSWER 1

You should have got something like this:

**Cri = {a, b, c, e}**

**V = {d}**

## IDEA

Now that we have some subsets, we can use them to analyze our data.

## SETS FROM DATA

**Superset**

A = {a, b, c, d, e}

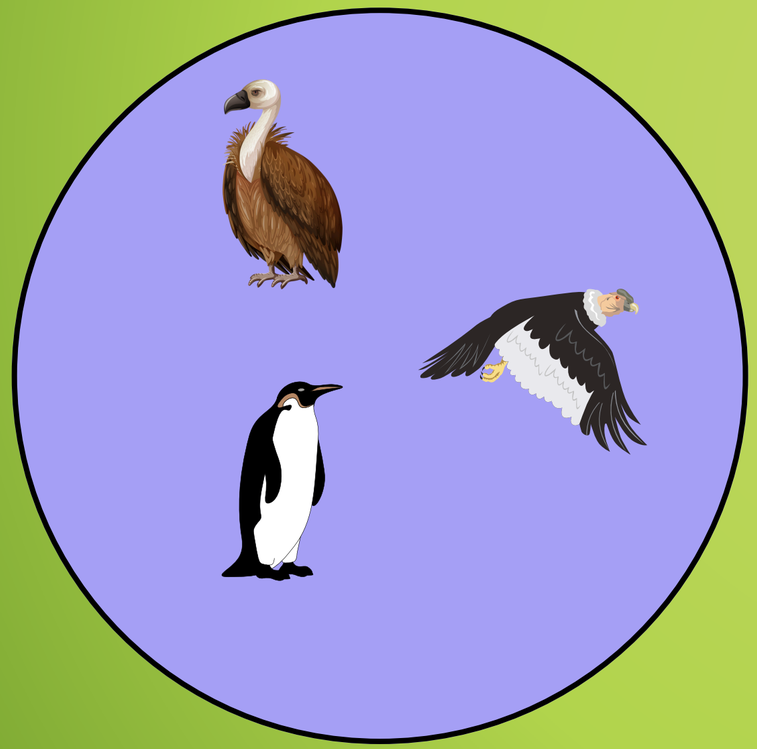
**Subsets**

H = {a}
C = {b, c, e}
O = {d}
L = {a, e}
W = {d}
Air = {b, c}
Cri = {a, b, c, e}
V = {d}

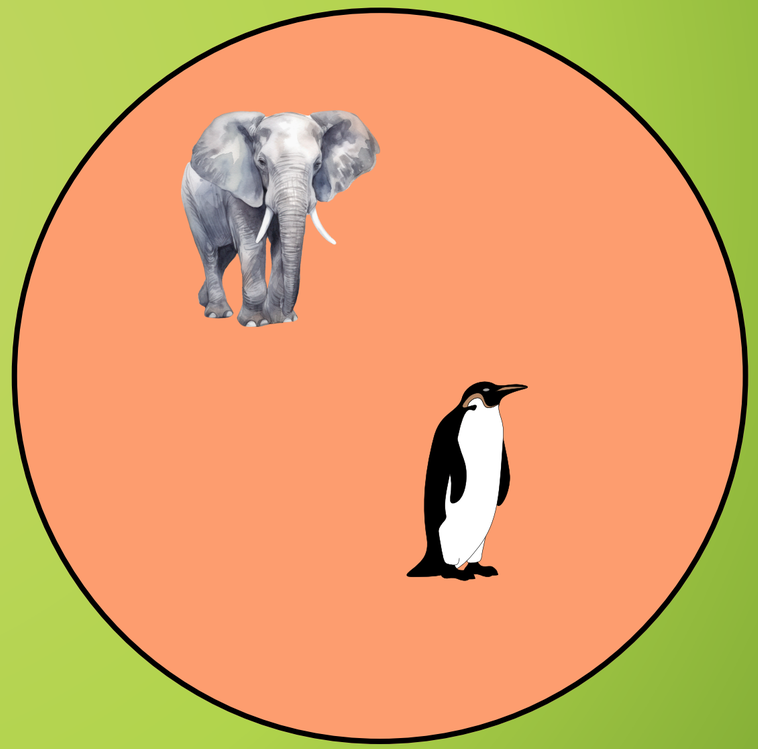
**DEFINITION**

The **union** of two sets, is a set containing every element in both sets.

## Carnivores (C)



## Land (L)



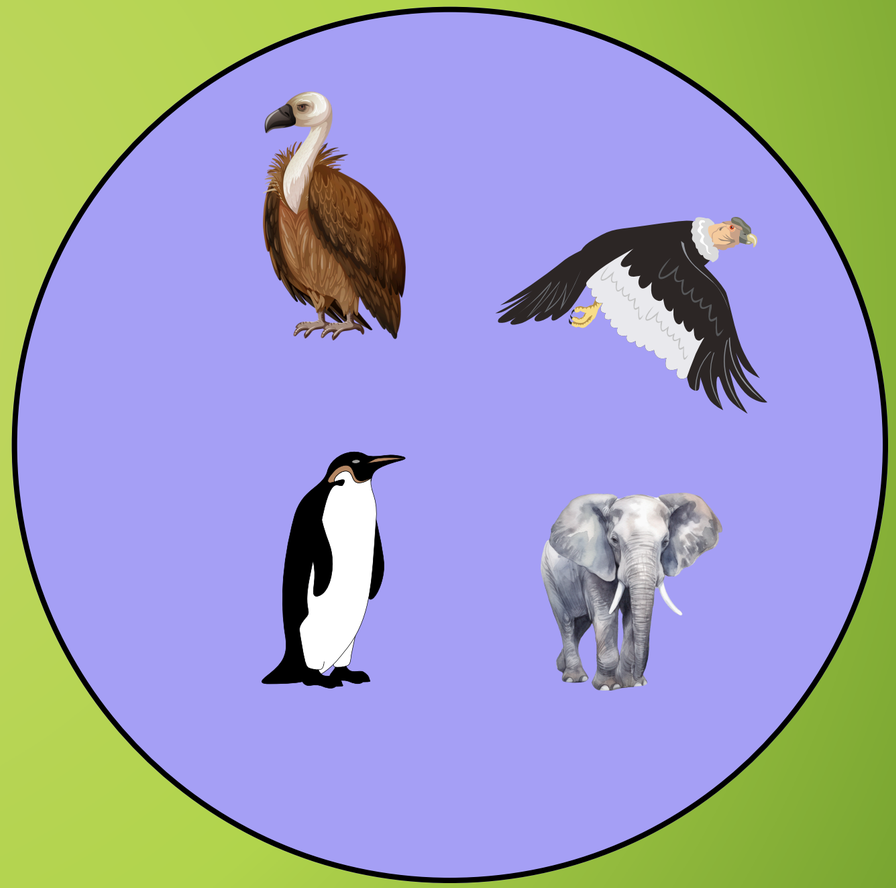
**DEFINITION**

The **union** of two sets, is a set containing every element in both sets.

**REMEMBER**

**We don’t repeat elements in sets!**

C U L =



## NOTATION

**C ∪ L** means “the union of set C and set L”

## QUESTION 2

If **L = {a, e}** and **Air = {b, c},** then

what is **L ∪ Air**?

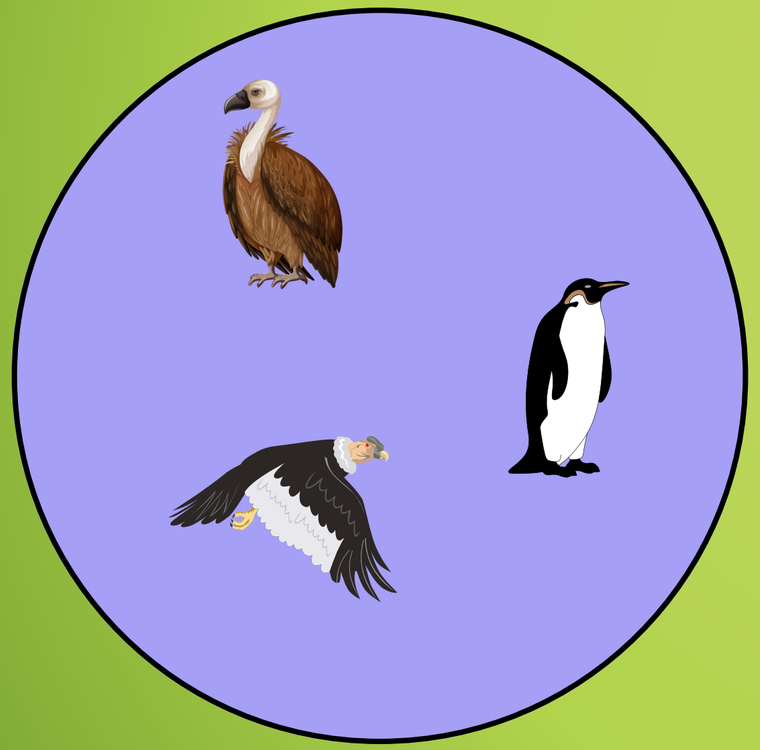
## ANSWER 2

$$L \cup Air = \{a, b, c, e\}$$

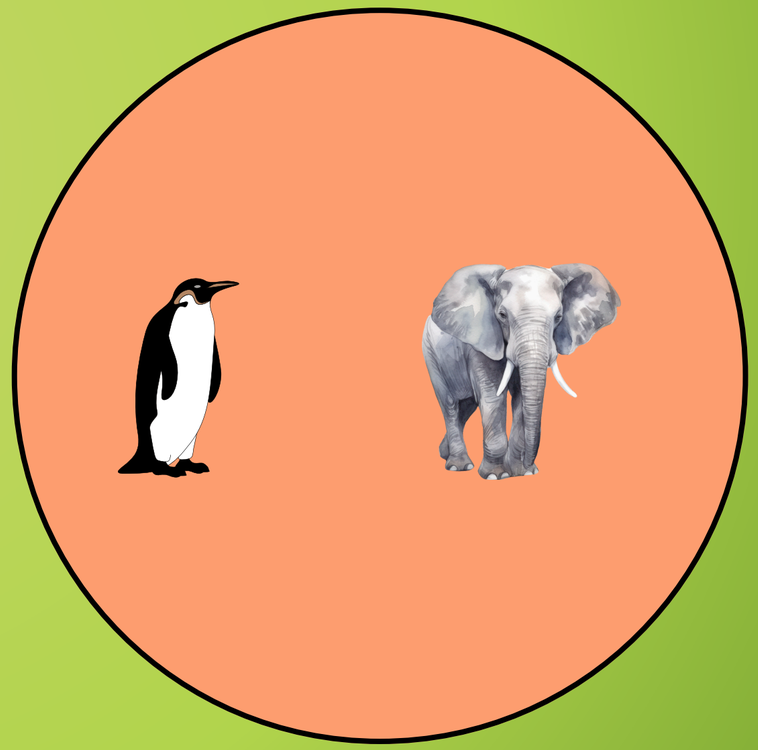
**DEFINITION**

The **intersection** of two sets, is a set containing every element that is a member of both sets.

## Carnivores (C)



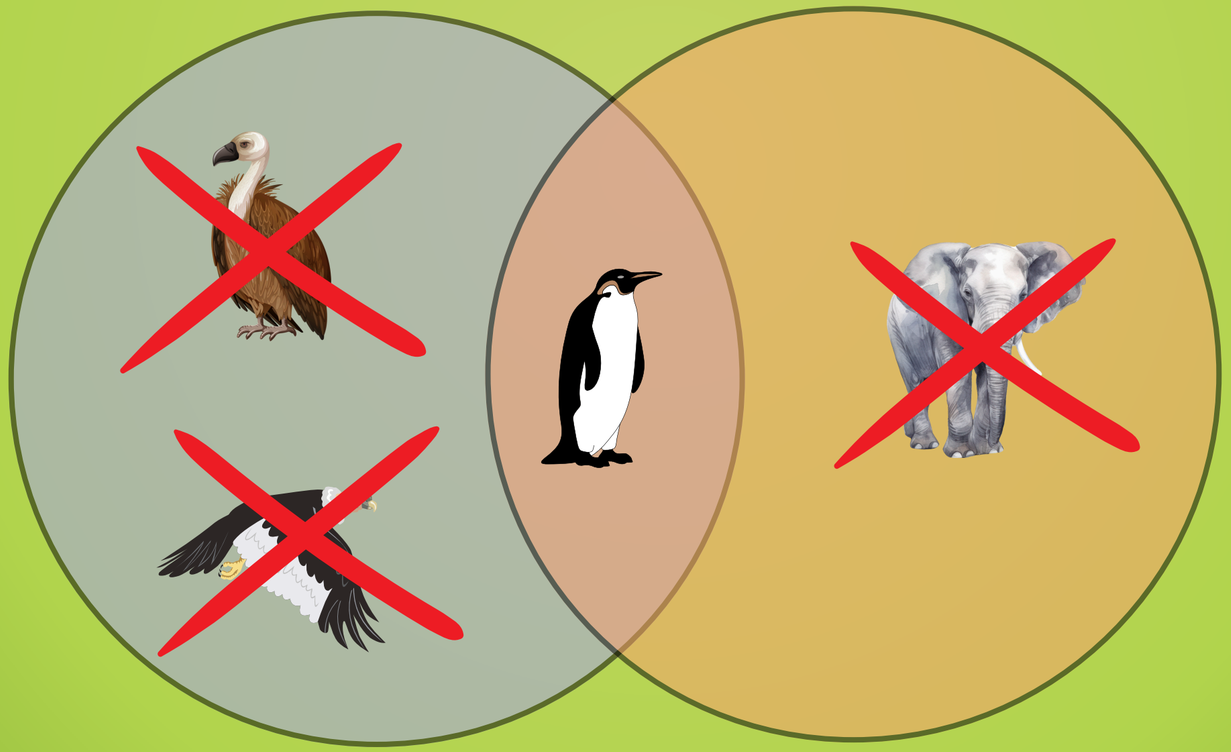
## Land (L)



**DEFINITION**

The intersection of two sets, is a set containing every element that is a member of both sets.

$C \cap L$



## NOTATION

**C ∩ L** means “the intersection of set C and set L”

## QUESTION 3

If **L = {a, e}** and **Cri = {a, b, c, e}**, then what is **L ∩ Cri**?

## ANSWER 3

$$L \cap Cri = \{a, e\} = L$$

## IDEA

Sometimes two sets won't have any elements in common at all.

This will result in the **empty set.**

If **H = {a}** and **O = {d}**, then **H ∩ O = {}.**

# IDEA

**We can use intersections and unions to calculate probabilities**

| C, Air | Is a carnivore | Is not a carnivore | Row Totals |
|---|---|---|---|
| **Lives in the air** | **Air** $\cap$ **C** | (**Air** $\cap$ **H**) $\cup$ (**Air** $\cap$ **O**) | |
| **Does not live in the air** | (**C** $\cap$ **L**) $\cup$ (**C** $\cap$ **W**) | Everything not in **C** or **Air** | |
| **Column Totals** | | | |

# IDEA

**We can use intersections and unions to calculate probabilities**

| C, Air | Is a carnivore | Is not a carnivore | Row Totals |
|---|---|---|---|
| **Lives in the air** | **{b, c}** | (**Air** ∩ **H**) ∪ (**Air** ∩ **O**) | |
| **Does not live in the air** | (**C** ∩ **L**) ∪ (**C** ∩ **W**) | Everything not in **C** or **Air** | |
| **Column Totals** | | | |

# IDEA

**We can use intersections and unions to calculate probabilities**

| C, Air | Is a carnivore | Is not a carnivore | Row Totals |
|---|---|---|---|
| **Lives in the air** | {b, c} | **{}∪{}** | |
| **Does not live in the air** | (**C**∩**L**)∪(**C**∩**W**) | Everything not in **C** or **Air** | |
| **Column Totals** | | | |

# IDEA

**We can use intersections and unions to calculate probabilities**

| C, Air | Is a carnivore | Is not a carnivore | Row Totals |
|---|---|---|---|
| **Lives in the air** | {b, c} | **{}** | |
| **Does not live in the air** | (**C**∩**L**)∪(**C**∩**W**) | Everything not in **C** or **Air** | |
| **Column Totals** | | | |

# IDEA

**We can use intersections and unions to calculate probabilities**

| C, Air | Is a carnivore | Is not a carnivore | Row Totals |
|---|---|---|---|
| **Lives in the air** | {b, c} | {} | |
| **Does not live in the air** | **{e}∪{}** | Everything not in **C** or **Air** | |
| **Column Totals** | | | |

# IDEA

**We can use intersections and unions to calculate probabilities**

| C, Air | Is a carnivore | Is not a carnivore | Row Totals |
|---|---|---|---|
| **Lives in the air** | {b, c} | {} | |
| **Does not live in the air** | **{e}** | Everything not in **C** or **Air** | |
| **Column Totals** | | | |

# IDEA

**We can use intersections and unions to calculate probabilities**

| C, Air | Is a carnivore | Is not a carnivore | Row Totals |
|---|---|---|---|
| **Lives in the air** | {b, c} | {} | |
| **Does not live in the air** | {e} | Everything not in **C** or **Air** | |
| **Column Totals** | | | |

# IDEA

**We can use intersections and unions to calculate probabilities**

| C, Air | Is a carnivore | Is not a carnivore | Row Totals |
|---|---|---|---|
| **Lives in the air** | {b, c} | {} | |
| **Does not live in the air** | {e} | **{a, d}** | |
| **Column Totals** | | | |

# IDEA

**We can use intersections and unions to calculate probabilities**

| C, Air | Is a carnivore | Is not a carnivore | Row Totals |
|---|---|---|---|
| **Lives in the air** | {b, c} | {} | |
| **Does not live in the air** | {e} | {a, d} | |
| **Column Totals** | **|{b, c}| + |{e}|** | **|{}| + |{a, d}|** | |

# IDEA

**We can use intersections and unions to calculate probabilities**

| C, Air | Is a carnivore | Is not a carnivore | Row Totals |
|---|---|---|---|
| **Lives in the air** | {b, c} | {} | |
| **Does not live in the air** | {e} | {a, d} | |
| **Column Totals** | **3** | **2** | |

# IDEA

**We can use intersections and unions to calculate probabilities**

| C, Air | Is a carnivore | Is not a carnivore | Row Totals |
|---|---|---|---|
| **Lives in the air** | {b, c} | {} | **\|{b, c}\| + \|{}\|** |
| **Does not live in the air** | {e} | {a, d} | **\|{e}\| + \|{a, d}\|** |
| **Column Totals** | 3 | 2 | **3 + 2** |

# IDEA

**We can use intersections and unions to calculate probabilities**

| C, Air | Is a carnivore | Is not a carnivore | Row Totals |
|---|---|---|---|
| **Lives in the air** | {b, c} | {} | **2** |
| **Does not live in the air** | {e} | {a, d} | **3** |
| **Column Totals** | 3 | 2 | **5** |

# IDEA

p(animal is a carnivore given that it doesn't live in the air) = $\frac{1}{3}$ = 0.3333

| C, Air | Is a carnivore | Is not a carnivore | Row Totals |
|---|---|---|---|
| **Lives in the air** | 2 | 0 | 2 |
| **Does not live in the air** | 1 | 2 | 3 |
| **Column Totals** | 3 | 2 | 5 |

# NOTATION

**p(animal is a carnivore given that it doesn’t live in the air)**

can be written as

**$p(x \in C \mid x \notin Air)$**

$\in$ means **“is a member of”**

$\notin$ means **“isn’t a member of”**

# QUESTION 4

What is $p(x \in C \mid x \in Air)$?

| C, Air | Is a carnivore | Is not a carnivore | Row Totals |
|---|---|---|---|
| Lives in the air | 2 | 0 | 2 |
| Does not live in the air | 1 | 2 | 3 |
| Column Totals | 3 | 2 | 5 |

# ANSWER 4

What is $p(x \in C \mid x \in Air) = \frac{2}{2} = 1$

| C, Air | Is a carnivore | Is not a carnivore | Row Totals |
|---|---|---|---|
| Lives in the air | 2 | 0 | 2 |
| Does not live in the air | 1 | 2 | 3 |
| Column Totals | 3 | 2 | 5 |

# QUESTION 5

What is p(x ∉ C | x ∉ Air)?

| C, Air | Is a carnivore | Is not a carnivore | Row Totals |
|---|---|---|---|
| Lives in the air | 2 | 0 | 2 |
| Does not live in the air | 1 | 2 | 3 |
| Column Totals | 3 | 2 | 5 |

# ANSWER 5

What is $p(x \notin C \mid x \notin Air) = \frac{2}{3} = 0.6667$

| C, Air | Is a carnivore | Is not a carnivore | Row Totals |
|---|---|---|---|
| **Lives in the air** | 2 | 0 | 2 |
| **Does not live in the air** | 1 | 2 | 3 |
| **Column Totals** | 3 | 2 | 5 |

# IDEA

For any two sets **X**, **Y**:

$$\mathbf{p}(x \in \mathbf{X} \mid x \in \mathbf{Y}) = \frac{|\mathbf{X} \cap \mathbf{Y}|}{|\mathbf{Y}|}$$

# RECAP

The **union** between two sets **A** and **B** is a set containing all elements that are members of A **or** B.

The **intersection** between two sets **A** and **B** is a set containing all elements that are members of A **and** B.

Intersections and unions can be used to calculate probabilities.