A Conceptual and Practical Introduction to Occupancy Models

Hanna Jackson

Les Ecols

October 16, 2024

Occupancy Models

- **Occupancy** is **whether** a species is present at a site or not

Site 1  Occupied

Site 2  Occupied

Site 3  Occupied

Site 4 Not occupied

Occupancy Models

- **Occupancy** is **whether** a species is present at a site or not
- Occupancy models infer **where** a species is found and **why** it is found there

Site 1  Occupied

Site 2  Occupied

Site 3  Occupied

Site 4 Not occupied

Occupancy as a Stochastic Process

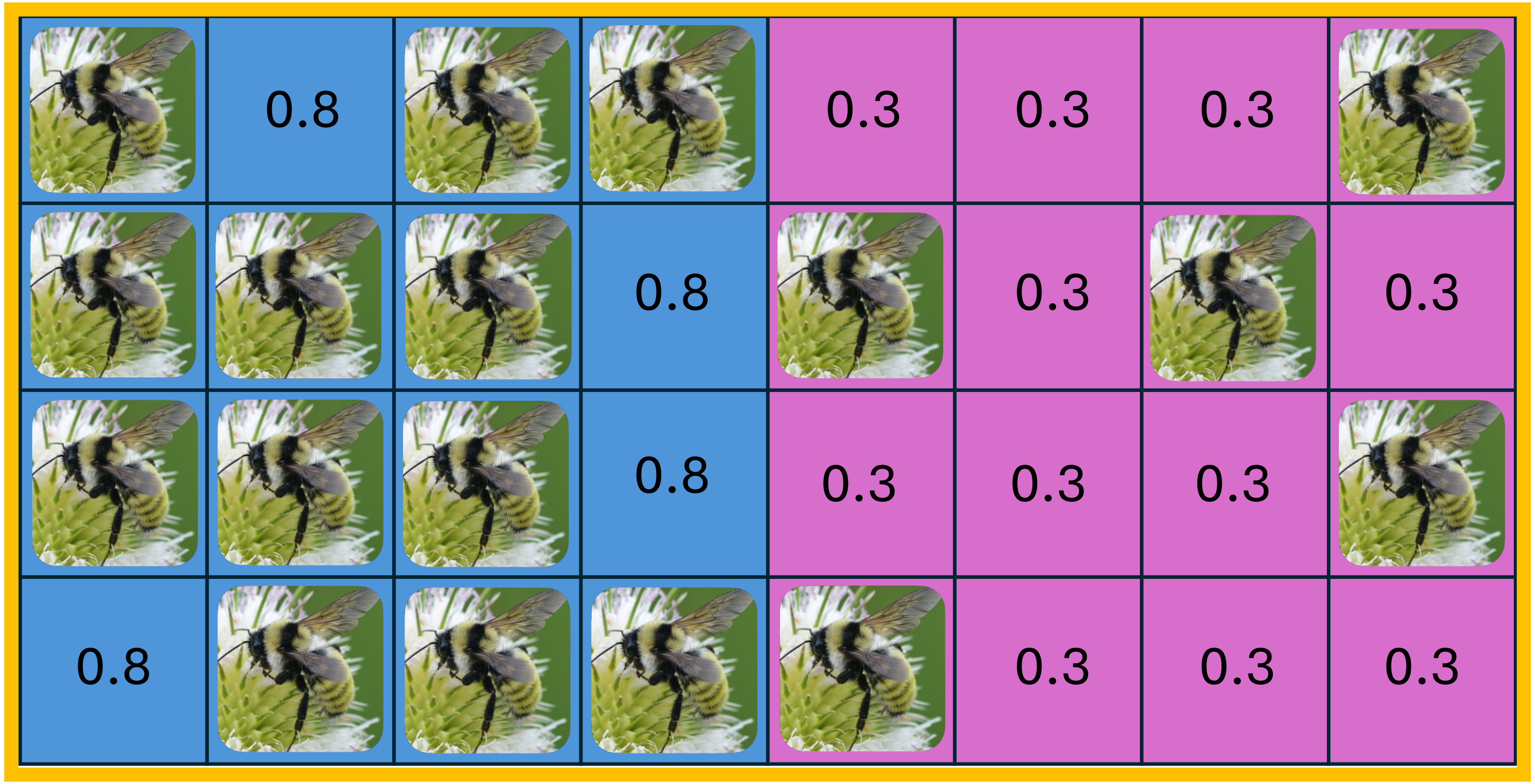
We assume that each site has a characteristic **occupancy probability**

0.8	0.8	0.8	0.8	0.3	0.3	0.3	0.3
0.8	0.8	0.8	0.8	0.3	0.3	0.3	0.3
0.8	0.8	0.8	0.8	0.3	0.3	0.3	0.3
0.8	0.8	0.8	0.8	0.3	0.3	0.3	0.3

Occupancy as a Stochastic Process

We assume that each site has a characteristic **occupancy probability**

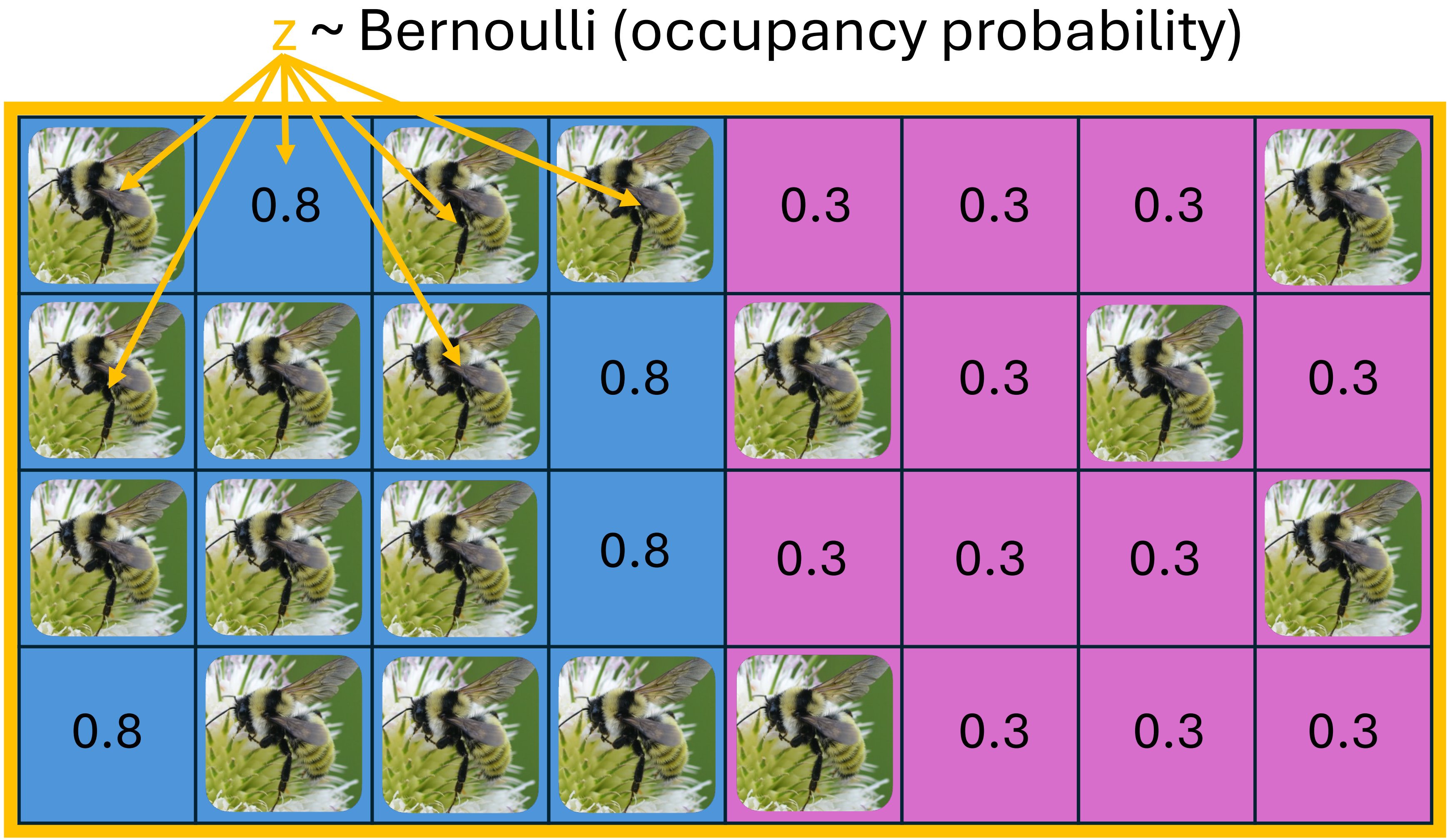
And that the sites are occupied with that probability



Occupancy as a Stochastic Process

We assume that each site has a characteristic **occupancy probability**

And that the sites are occupied with that probability



Occupancy as a Stochastic Process

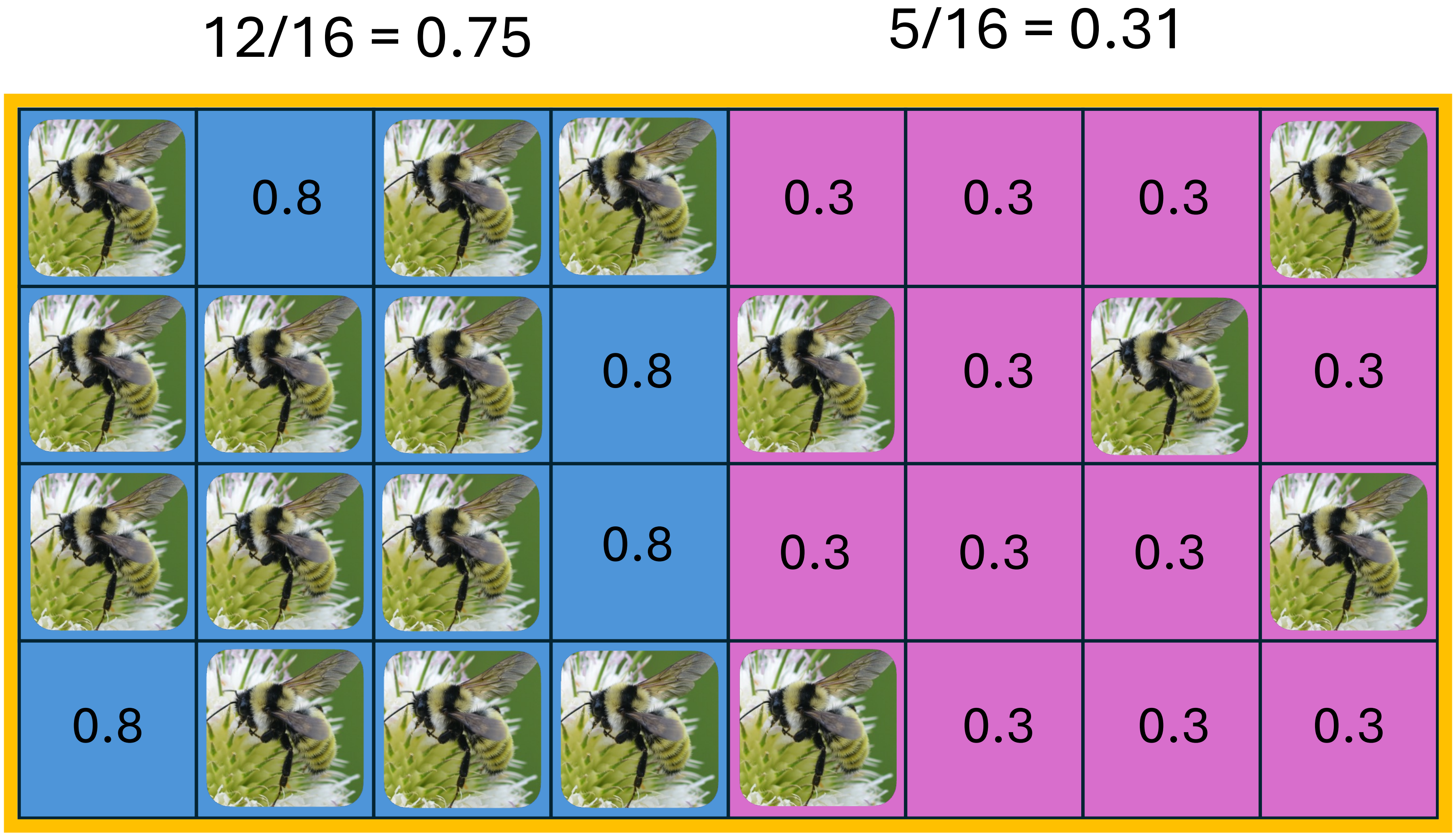
z ~ Coin Flip (occupancy probability)

We assume that each site has a characteristic **occupancy probability**

And that the sites are occupied with that probability

Occupancy as a Stochastic Process

So given a set of sites with known **occupancy**, we can get our best estimate of the **occupancy probability**



Occupancy as a Stochastic Process

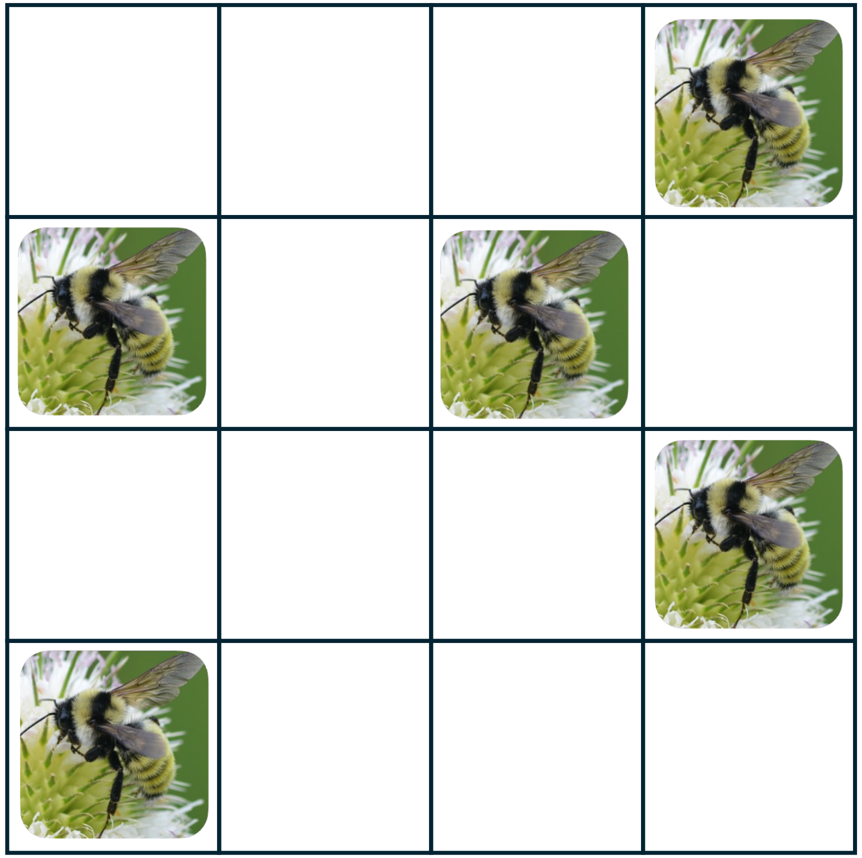
So given a set of sites with known **occupancy**, we can get our best estimate of the **occupancy probability**

But knowing the true occupancy of the site is hard!

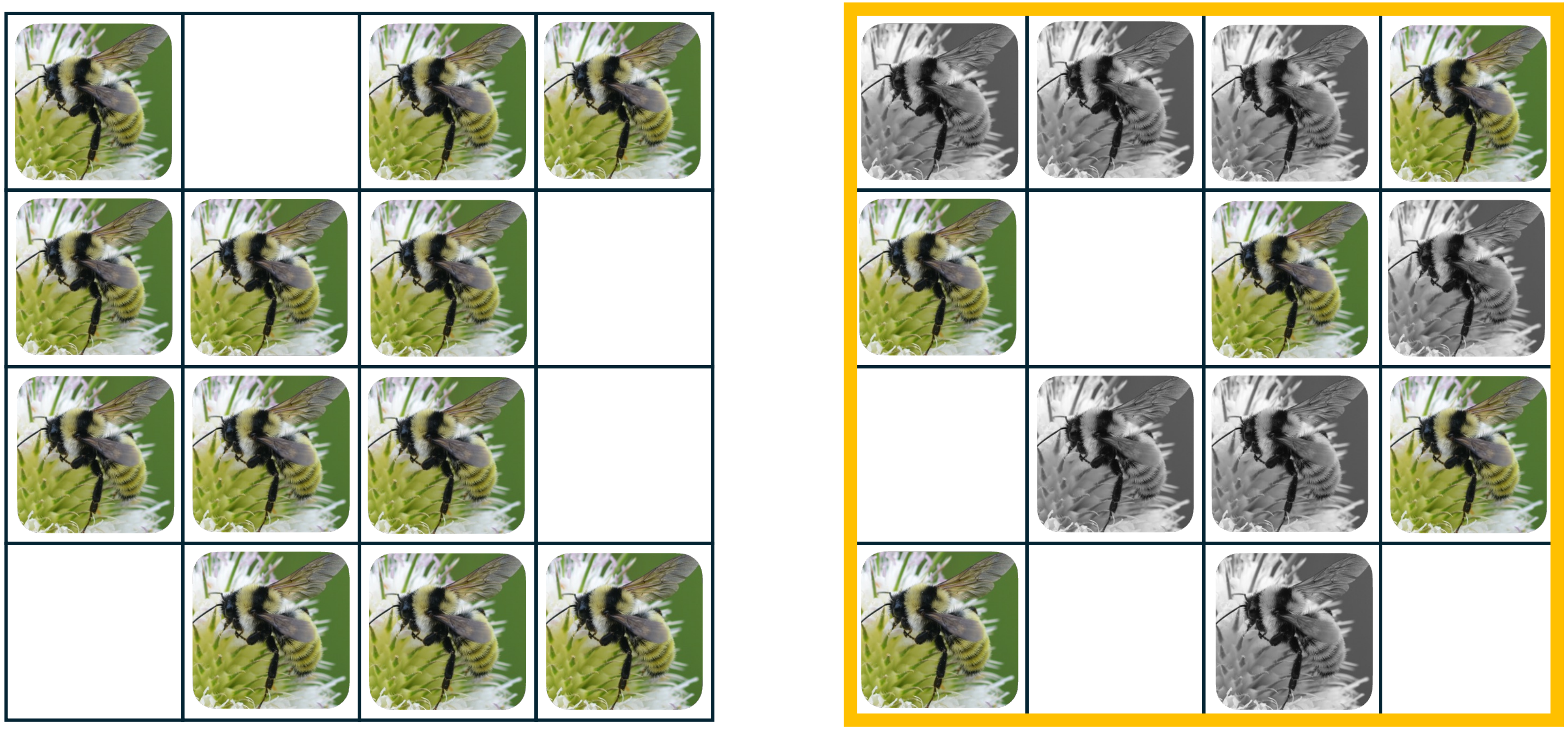
12/16 = 0.75

5/16 = 0.31

Finding the true occupancy of the sites is hard!

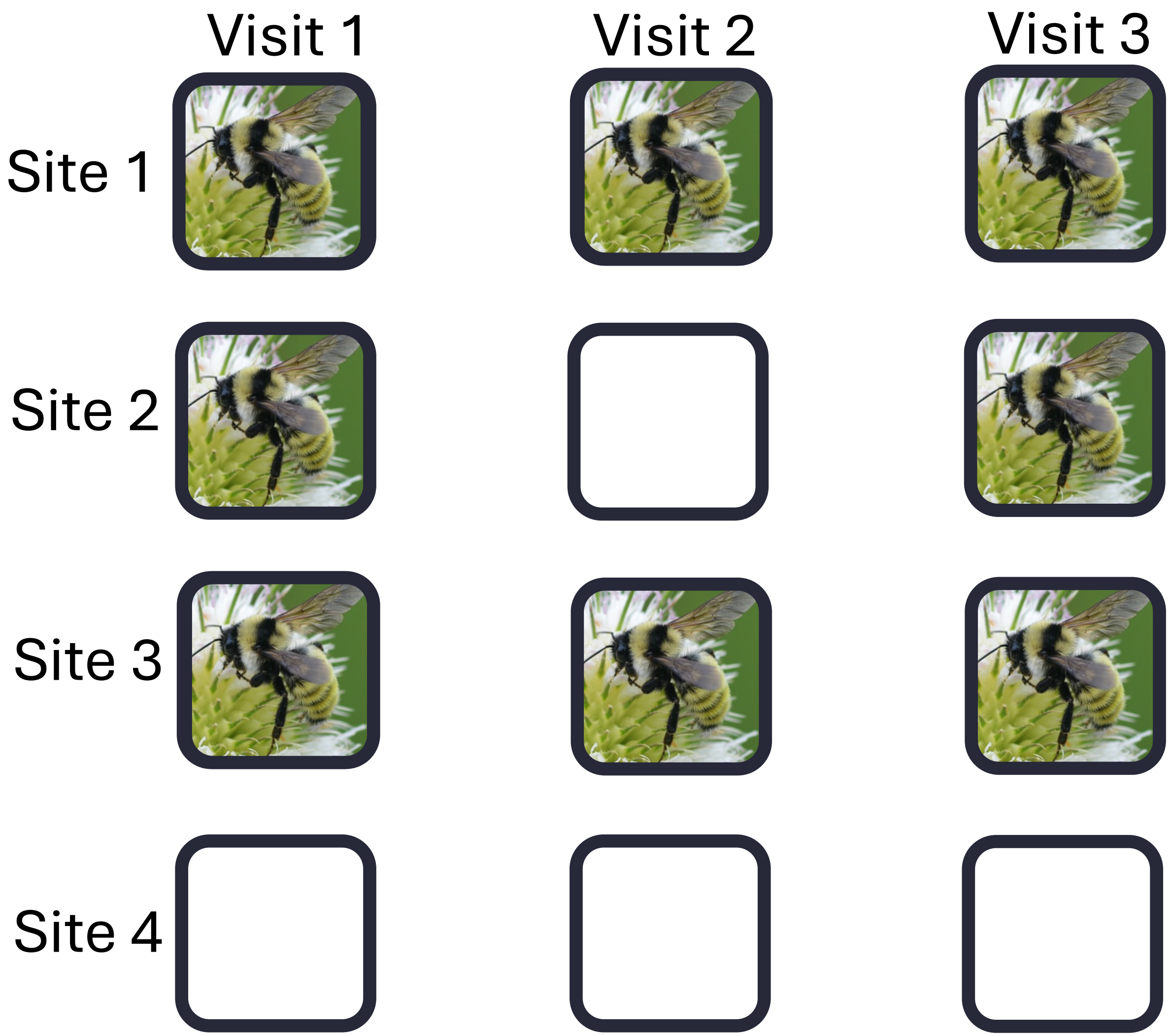


Finding the true occupancy of the sites is hard!



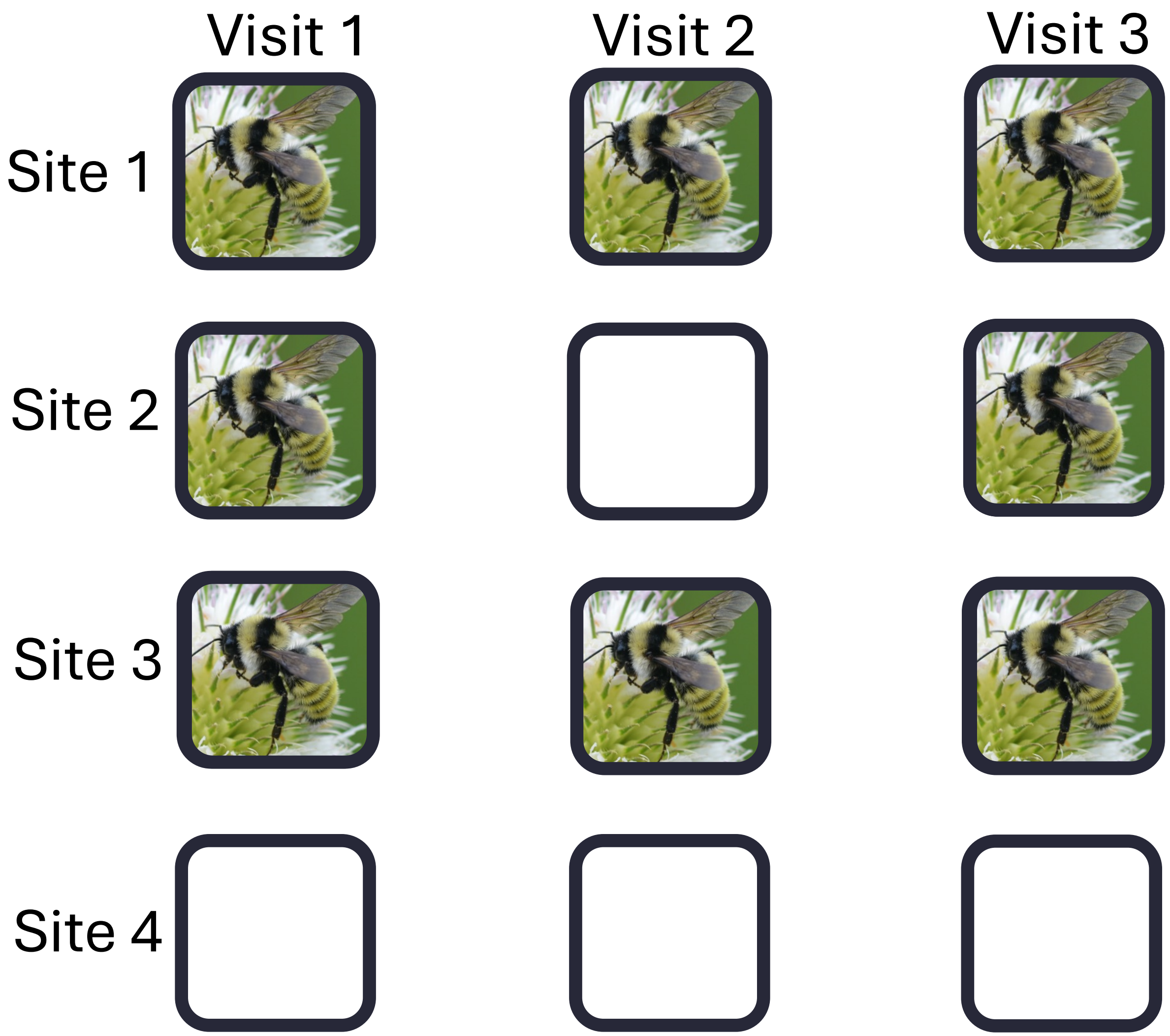
Occupancy Models

Occupancy Models



Occupancy Models

How is this structured data useful?



Occupancy Models

We **know** this species is at these sites

Occupancy Models

We **know** this species is at these sites

Absent from this site
or
did we miss it?

Occupancy Models

We appear to be **GOOD** at **finding bees** when they are present

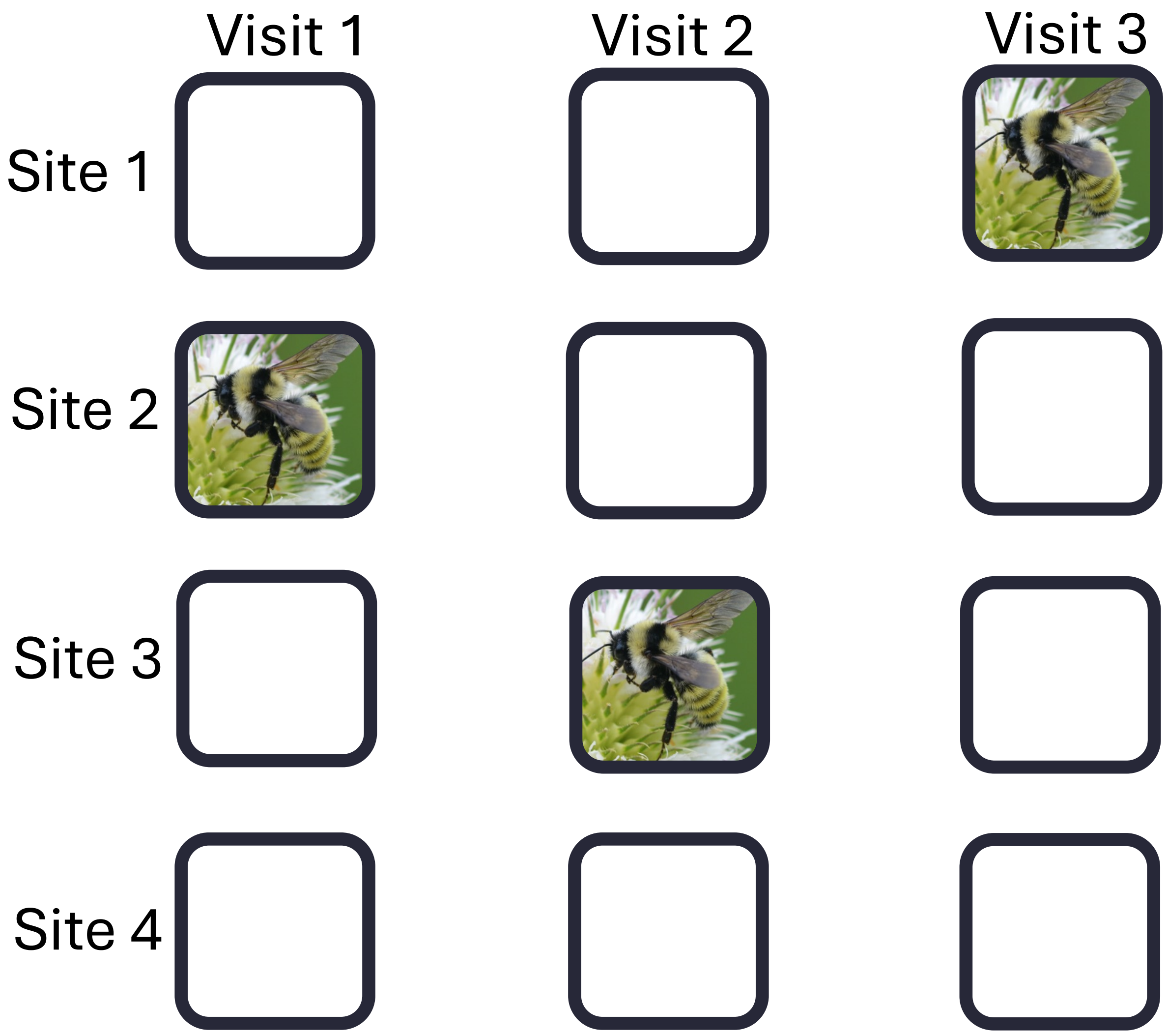
Absent from this site
or
did we miss it?

Occupancy Models

We appear to be **GOOD** at **finding bees** when they are present

High Detection Probability

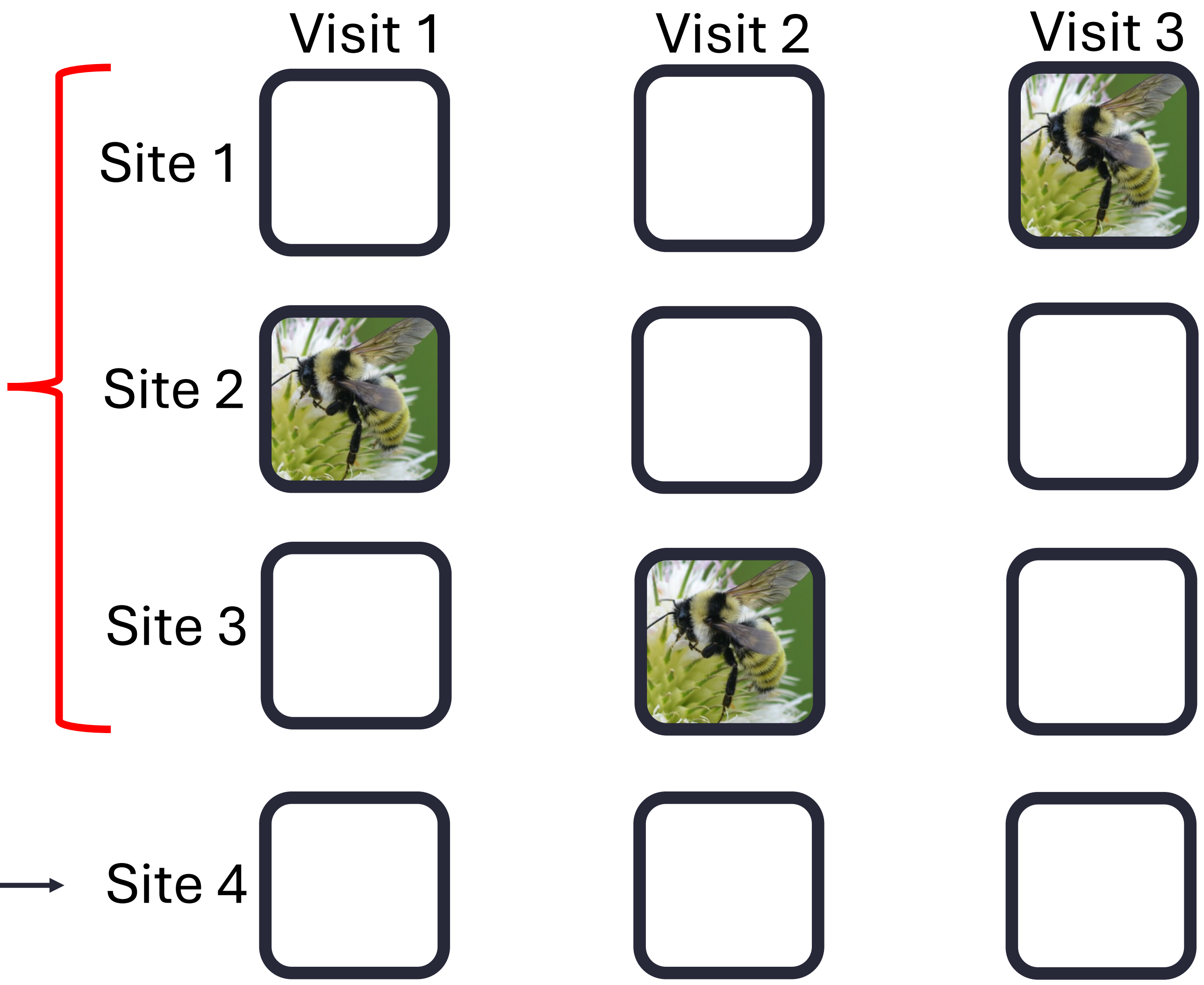
Occupancy Models



Absent from this site
or
did we miss it?

Occupancy Models

We appear to be **BAD at finding bees** when they are present



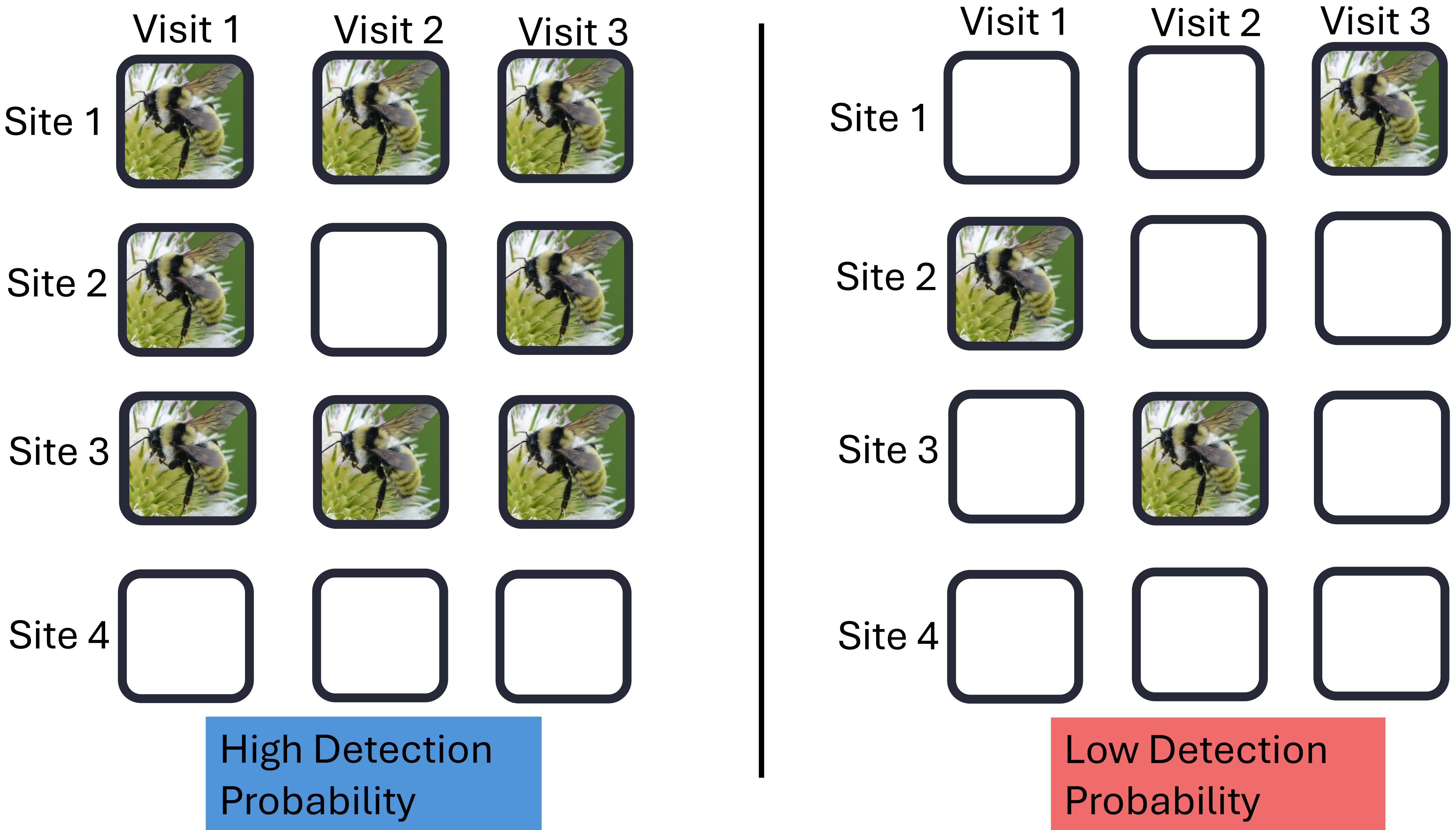
Absent from this site
or
did we miss it?

Occupancy Models

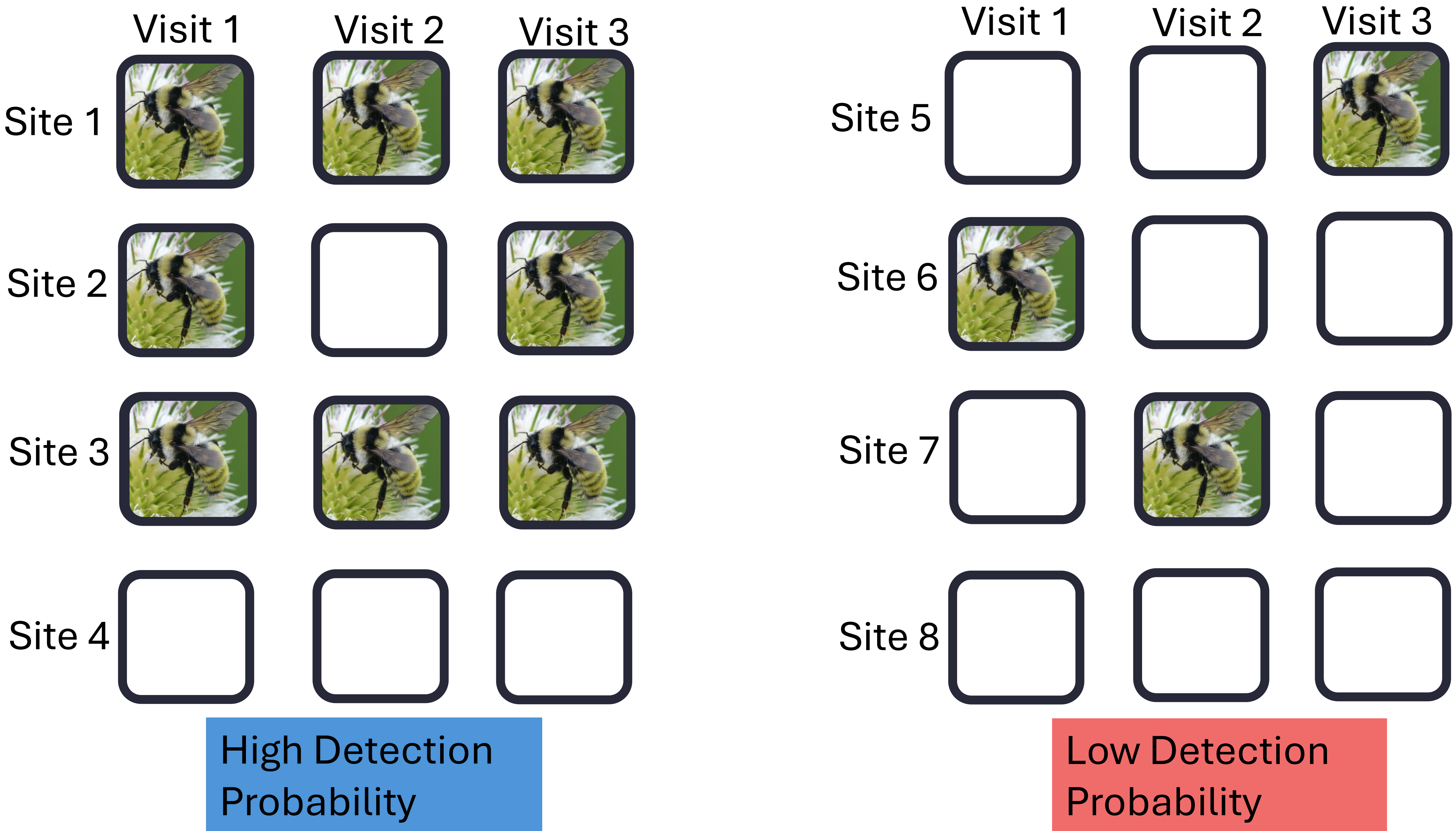
We appear to be **BAD at finding bees** when they are present

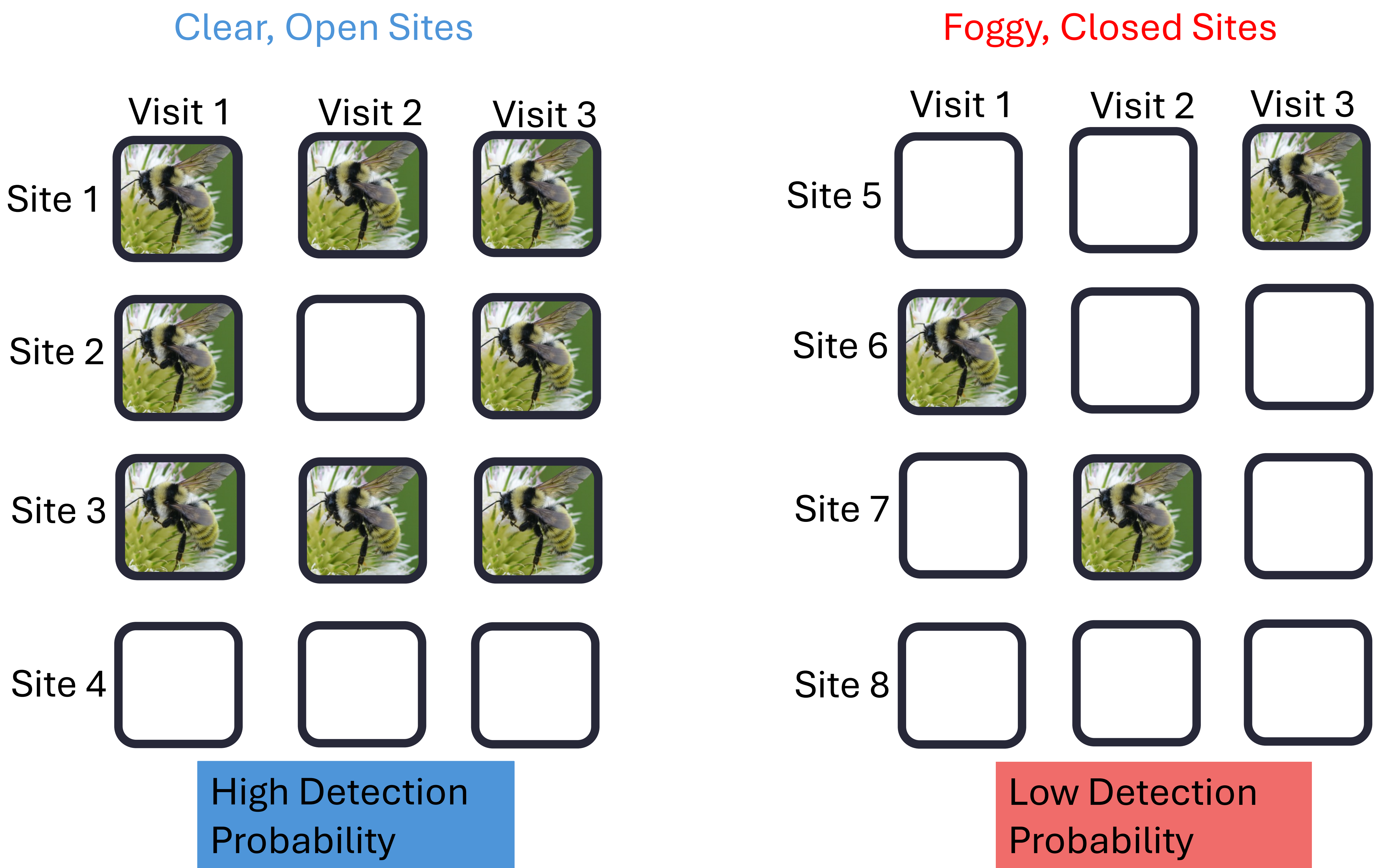
Low Detection Probability

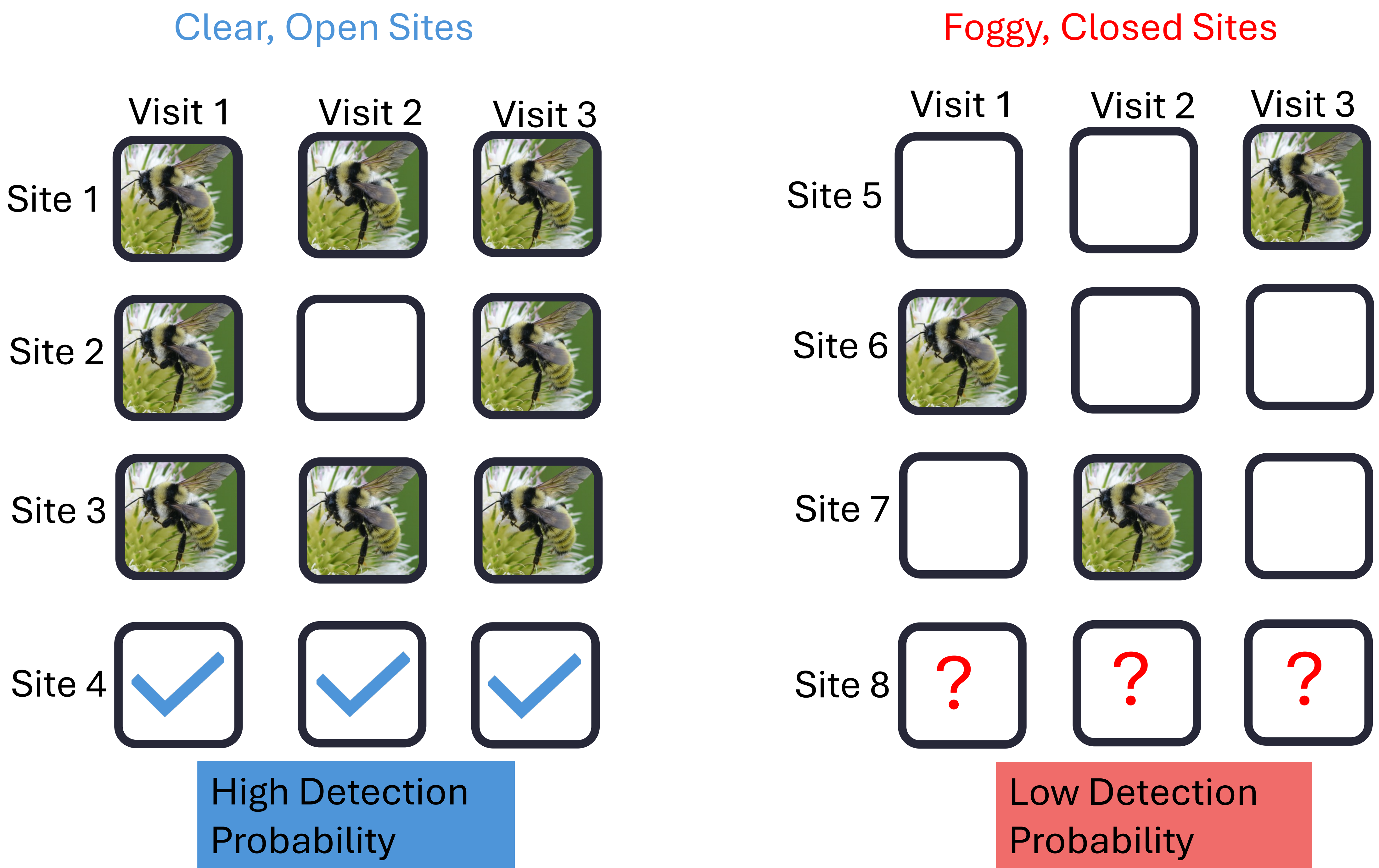
But what if instead of two separate scenarios...



… It was one study, with two kinds of sites




Clear, Open Sites
Foggy, Closed Sites
Visit 1
Visit 2
Visit 3
Visit 1
Visit 2
Visit 3
Site 1
Site 2
Site 3
Site 4
Site 5
Site 6
Site 7
Site 8
High Detection Probability
Low Detection Probability


Clear, Open Sites
Foggy, Closed Sites
Visit 1
Visit 2
Visit 3
Site 1
Site 2
Site 3
Site 4
Visit 1
Visit 2
Visit 3
Site 5
Site 6
Site 7
Site 8
?
?
?
High Detection Probability
Low Detection Probability

Takeaway:

Occupancy models:

- Use **repeat visits** to sites
- To quantify **detection probability**
- So that when we model **occupancy probability** we are **accounting for this uncertainty** in whether observed absences are true zeros or missed detections

Hierarchical Models

A "Hierarchical Model" is two GLMs that are connected to each other *

$$\text{logit}\left(\boxed{\begin{array}{c}\text{State Process:}\\ \text{Occupancy}\\ \text{Probability}\end{array}}\right) = \alpha 0 + \alpha 1 * X1 + \alpha 2 * X2 + \alpha 3 * X1 * X2$$

$$\text{logit}\left(\boxed{\begin{array}{c}\text{Observation Process:}\\ \text{Detection}\\ \text{Probability}\end{array}}\right) = \beta 0 + \beta 1 * X1 + \beta 2 * X2 + \beta 3 * X1 * X2$$

Hierarchical Models

A "Hierarchical Model" is two GLMs that are connected to each other *

$$\text{logit}\left(\boxed{\begin{array}{c}\text{State Process:}\\ \text{Occupancy}\\ \text{Probability}\end{array}}\right) = \psi0 + \psi1*X1 + \psi2*X2 + \psi3*X1*X2$$

$$\text{logit}\left(\boxed{\begin{array}{c}\text{Observation Process:}\\ \text{Detection}\\ \text{Probability}\end{array}}\right) = p0 + p1*X1 + p2*X2 + p3*X1*X2$$

But how does that work?

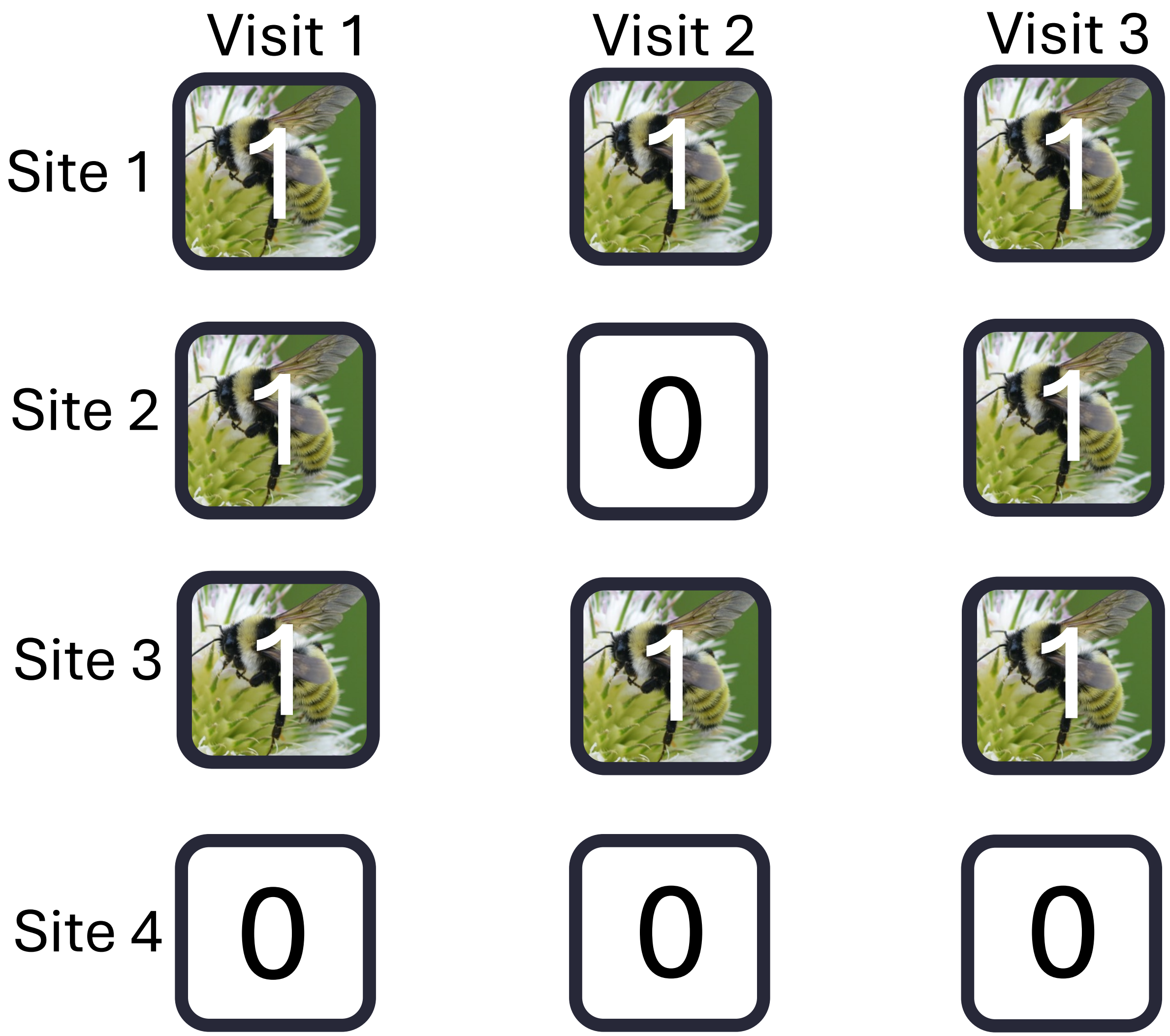
We appear to be **GOOD** at **finding bees** when they are present

High Detection Probability

Occupancy Models - Gears and Cogs

Each of these observations will be denoted X going forward

(I’ve coloured X green because they’re our field observations)

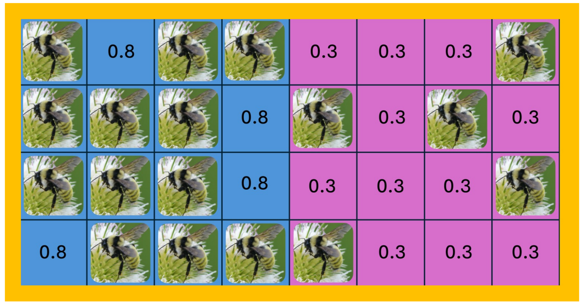


Occupancy Models - Gears and Cogs

$$\text{logit}\left(\boxed{\text{Occupancy Probability}}\right) = \psi.0 + \psi1 * X1 + \psi2 * X2 + \psi3 * X1 * X2$$

$$\text{logit}\left(\boxed{\text{Detection Probability}}\right) = p.0 + p1 * X1 + p2 * X2 + p3 * X1 * X2$$

Occupancy Models - Gears and Cogs

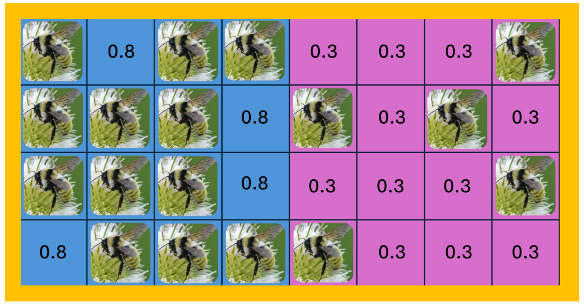


$$\text{logit}\left(\text{Occupancy Probability}\right) = \psi.0 + \psi1{*}X1 + \psi2{*}X2 + \psi3{*}X1{*}X2$$

$$\text{logit}\left(\text{Detection Probability}\right) = p.0 + p1{*}X1 + p2{*}X2 + p3{*}X1{*}X2$$

Occupancy Models - Gears and Cogs

Z ~ Coin Flip (Occupancy Probability)



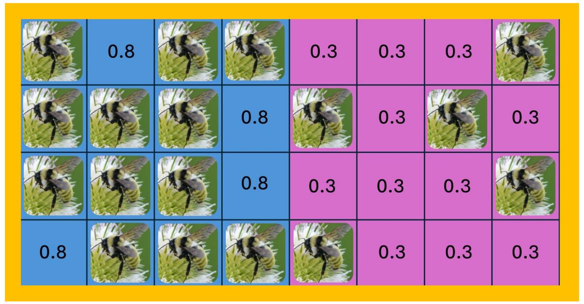
logit (Occupancy Probability) = ψ.0 + ψ1*X1 + ψ2*X2 + ψ3*X1*X2

logit (Detection Probability) = p.0 + p1*X1 + p2*X2 + p3*X1*X2

Occupancy Models - Gears and Cogs

Observations, X don't <u>just</u> depend on the probability the species is present

Z ~ Coin Flip (Occupancy Probability)



logit (Occupancy Probability) = ψ.0 + ψ1*X1 + ψ2*X2 + ψ3*X1*X2

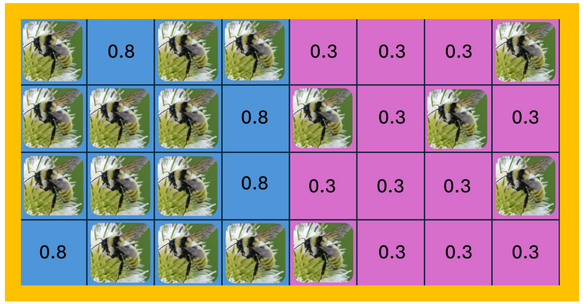
logit (Detection Probability) = p.0 + p1*X1 + p2*X2 + p3*X1*X2

Occupancy Models - Gears and Cogs

Observations, X don’t <u>just</u> depend on the probability the species is present

X ~ Coin Flip (Z * $\boxed{\text{Detection Probability}}$)

Z ~ Coin Flip ($\boxed{\text{Occupancy Probability}}$)



logit ($\boxed{\text{Occupancy Probability}}$) = $\psi.0 + \psi1*X1 + \psi2*X2 + \psi3*X1*X2$

logit ($\boxed{\text{Detection Probability}}$) = $p.0 + p1*X1 + p2*X2 + p3*X1*X2$

Occupancy Models - Gears and Cogs

Observations, X don’t just depend on the probability the species is present

X ~ Coin Flip (Z * Detection Probability)

Z ~ Coin Flip (Occupancy Probability)

At all sites where the species was found, Z must be 1

logit (Occupancy Probability) = ψ.0 + ψ1*X1 + ψ2*X2 + ψ3*X1*X2

logit (Detection Probability) = p.0 + p1*X1 + p2*X2 + p3*X1*X2

Occupancy Models - Gears and Cogs

At all sites where the species was found, Z must be 1

For one such site,

1 ~ Coin Flip (Occupancy Probability)

Visit 1

Visit 2

Visit 3

0

Occupancy Models - Gears and Cogs

At all sites where the species was found, Z must be 1

For one such site,

1 ~ Coin Flip (Occupancy Probability)

Visit 1 

~ Coin Flip (1 * Detection Probability)

Visit 2 

~ Coin Flip (1 * Detection Probability)

Visit 3 0 ~ Coin Flip (1 * Detection Probability)

Occupancy Models - Gears and Cogs

At all sites where the species was found, Z must be 1

For one such site,

1 ~ Coin Flip (Occupancy Probability)

Visit 1 

~ Coin Flip (1 * Detection Probability)

Visit 2 

~ Coin Flip (1 * Detection Probability)

Visit 3 0 ~ Coin Flip (1 * Detection Probability)

For two successes and one failure, the most likely detection probability is 0.66

Detection Probability = 0.66

Occupancy Models - Gears and Cogs

At all sites where the species was not found, Z could be 1 OR 0

1 ~ Coin Flip (Occupancy Probability)

0 ~ Coin Flip (Occupancy Probability)

Occupancy Models - Gears and Cogs

At all sites where the species was not found, Z could be 1 OR 0

1 ~ Coin Flip (Occupancy Probability)

0 ~ Coin Flip (Occupancy Probability)

0 ~ Coin Flip (1 * Detection Probability)

0 ~ Coin Flip (1 * Detection Probability)

0 ~ Coin Flip (1 * Detection Probability)

0 ~ Coin Flip (0 * Detection Probability)

0 ~ Coin Flip (0 * Detection Probability)

0 ~ Coin Flip (0 * Detection Probability)

Occupancy Models - Gears and Cogs

At all sites where the species was not found, Z could be 1 OR 0

1 ~ Coin Flip (Occupancy Probability)

0 ~ Coin Flip (Occupancy Probability)

0 ~ Coin Flip (1 * Detection Probability)

0 ~ Coin Flip (0 * Detection Probability)

0 ~ Coin Flip (

Both sides here are consistent with the data (all zeros)

lip (0 * Detection Probability)

0 ~ Coin Flip (1 * Detection Probability)

0 ~ Coin Flip (0 * Detection Probability)

Hierarchical* Models are weird, but cool!

- Normally, your ’y’ is something that you have observed or calculated from your observations of the world
- Here, things are different
  1. We have TWO y’s

     Occupancy Probability

     Detection Probability

  2. AND neither of them are something we’ve directly measured

     Instead, we have observations, **X**, that depend on both of those quantities AND we have Z, the true occupancy state that we are also estimating
- But, their structure allows for us to systematically correct for biases in the data