Multi-Species Occupancy Models: Graphing Output

- Arielle Parsons
- North Carolina State University
- North Carolina Museum of Natural Sciences

Outline

- A review of marginal and conditional probabilities
- Calculating the natural parameters (f's) and occupancy probabilities
- Manipulating MCMC output in R
- An intro to ggplot2
- Graphing

Marginal Probabilities

- The marginal distribution of Y is its distribution averaged over all possible values of X

- In the context of this model, it is the occupancy of each individual species regardless of presence and absence of the other species

- Can be calculated along a covariate gradient
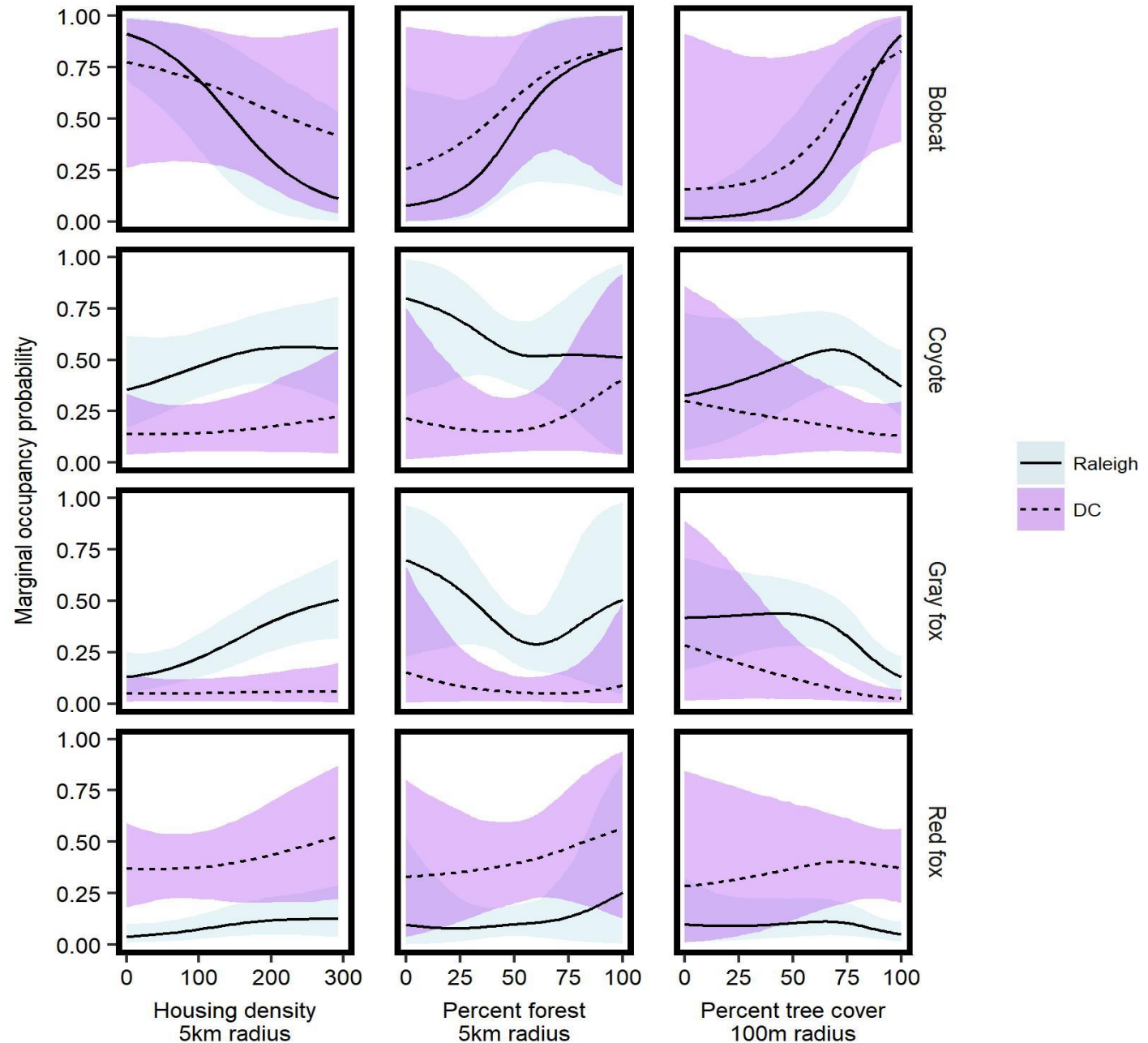
Conditional Probabilities

- The distribution of Y at a particular value of X

- In the context of this model, that means the probability of occurrence of a given species conditional on the presence/ absence of a different species

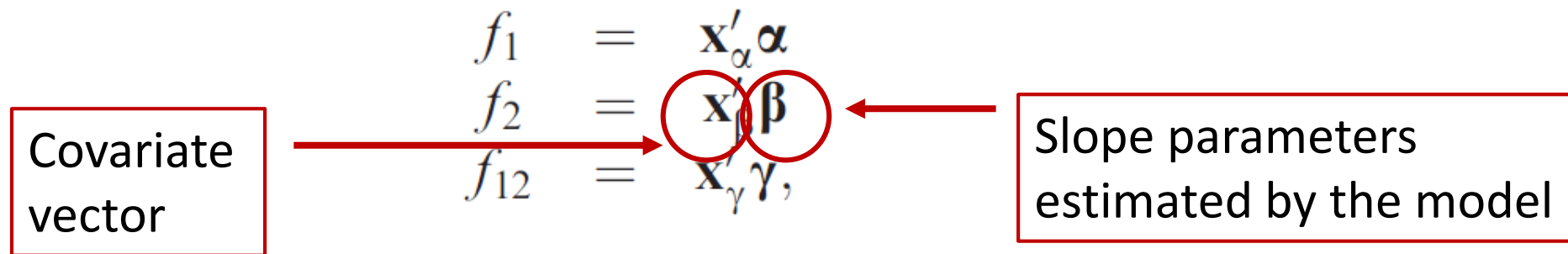- Can be calculated along a covariate gradient
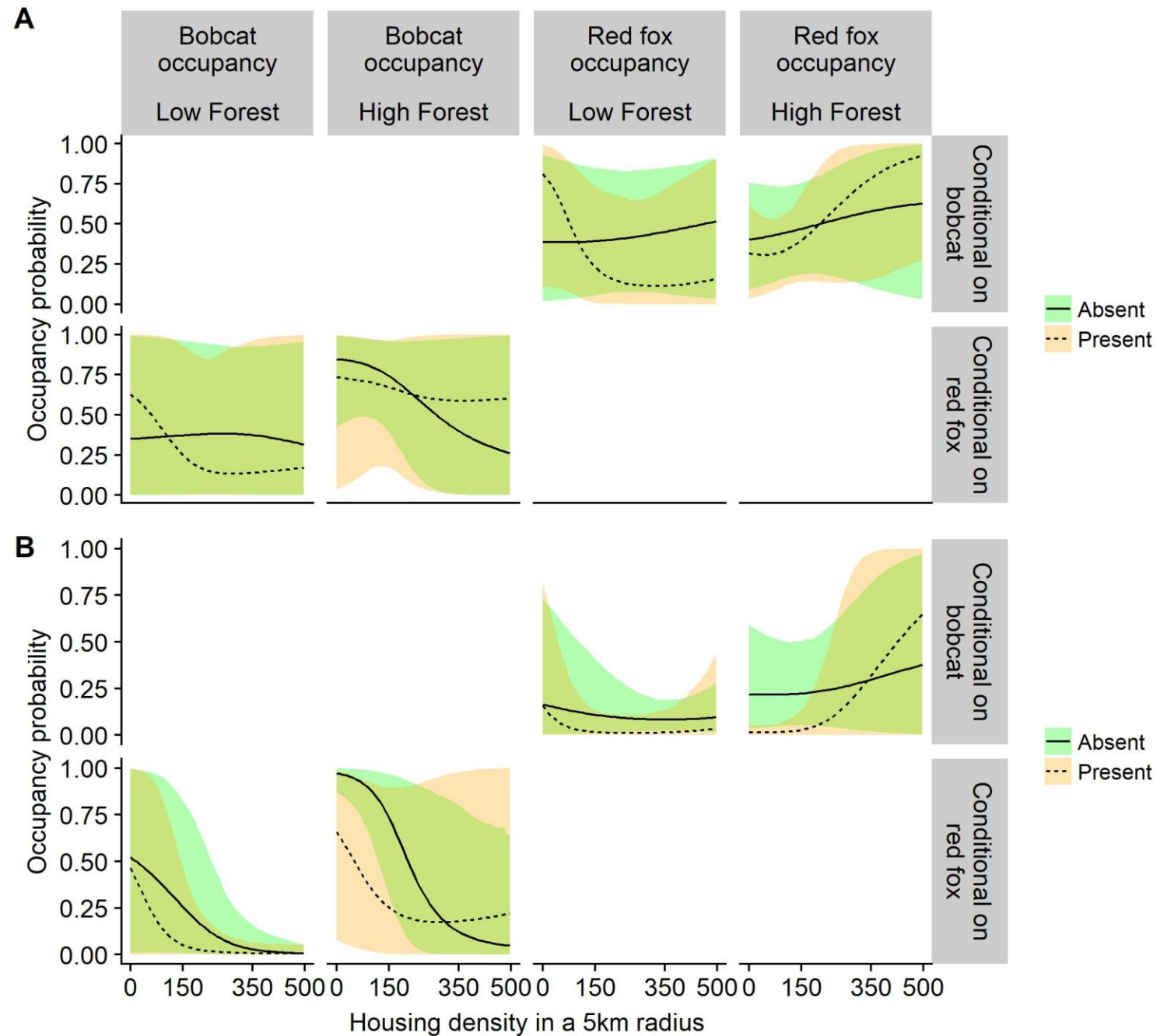
The natural parameters (*f* parameters)

- The *f* parameters in the multi-species model can be defined as linear functions of covariates (in a 2-species example):

$$
\begin{aligned}
f_1 &= \mathbf{x}'_\alpha \boldsymbol{\alpha} \\
f_2 &= \mathbf{x}'_\beta \boldsymbol{\beta} \\
f_{12} &= \mathbf{x}'_\gamma \boldsymbol{\gamma},
\end{aligned}
$$

The natural parameters (*f* parameters)

- The *f* parameters in the multi-species model can be defined as linear functions of covariates (in a 2-species example):

$$f_1 = \mathbf{x}'_\alpha \boldsymbol{\alpha}$$
$$f_2 = \mathbf{x}'_\beta \boldsymbol{\beta}$$
$$f_{12} = \mathbf{x}'_\gamma \boldsymbol{\gamma},$$

Covariate vector

Slope parameters estimated by the model

$$\beta_0 * 1 + \beta_1 * Dist5km + \beta_2 * HDens5km$$

The natural parameters (*f* parameters)

- The natural parameters are likewise defined as log odds that a species occupies a site:

$$f_1 = \log\left(\frac{\psi_{10}}{\psi_{00}}\right)$$

$$f_2 = \log\left(\frac{\psi_{01}}{\psi_{00}}\right)$$

$$f_{12} = \log\left(\frac{\psi_{11}\,\psi_{00}}{\psi_{01}\,\psi_{10}}\right)$$

Getting occupancy probabilities

- So we can use the *f* parameters to get conditional and marginal occupancy probabilities using a multinomial logit link

$$\psi_{11} = \frac{\exp(f_1 + f_2 + f_{12})}{1 + \exp(f_1) + \exp(f_2) + \exp(f_1 + f_2 + f_{12})},$$

$$\psi_{10} = \frac{\exp(f_1)}{1 + \exp(f_1) + \exp(f_2) + \exp(f_1 + f_2 + f_{12})},$$

$$\psi_{01} = \frac{\exp(f_2)}{1 + \exp(f_1) + \exp(f_2) + \exp(f_1 + f_2 + f_{12})},$$

$$\psi_{00} = \frac{1}{1 + \exp(f_1) + \exp(f_2) + \exp(f_1 + f_2 + f_{12})}.$$

Marginal occupancy probabilities:

Sum over all capture histories where species x was present, regardless of whether other species were also present

$$\psi_{11} = \frac{\exp(f_1+f_2+f_{12})}{1+\exp(f_1)+\exp(f_2)+\exp(f_1+f_2+f_{12})},$$

$$\psi_{10} = \frac{\exp(f_1)}{1+\exp(f_1)+\exp(f_2)+\exp(f_1+f_2+f_{12})},$$

$$\psi_{01} = \frac{\exp(f_2)}{1+\exp(f_1)+\exp(f_2)+\exp(f_1+f_2+f_{12})},$$

$$\psi_{00} = \frac{1}{1+\exp(f_1)+\exp(f_2)+\exp(f_1+f_2+f_{12})}.$$

Marginal occupancy Species 1: $\psi_{11} + \psi_{10}$

Marginal occupancy Species 2: $\psi_{11} + \psi_{01}$
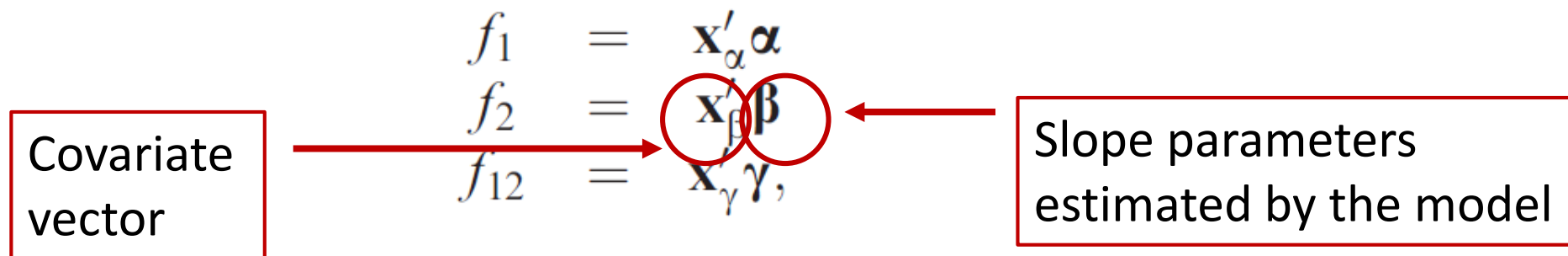
Conditional occupancy probabilities:

$$\psi_{11} = \frac{\exp(f_1+f_2+f_{12})}{1+\exp(f_1)+\exp(f_2)+\exp(f_1+f_2+f_{12})},$$

$$\psi_{10} = \frac{\exp(f_1)}{1+\exp(f_1)+\exp(f_2)+\exp(f_1+f_2+f_{12})},$$

$$\psi_{01} = \frac{\exp(f_2)}{1+\exp(f_1)+\exp(f_2)+\exp(f_1+f_2+f_{12})},$$

$$\psi_{00} = \frac{1}{1+\exp(f_1)+\exp(f_2)+\exp(f_1+f_2+f_{12})}.$$

Occupancy of Species 1 *in the absence* of Species 2 : $\psi_{10}/\psi_{10}+\psi_{00}$

Occupancy of Species 1 *in the presence* of Species 2: $\psi_{11}/\psi_{11}+\psi_{01}$

Manipulating MCMC output in R

- Step 1: Load the MCMC object
- Step 2: Extract the beta coefficients
- Step 3: Read in the original covariates used to fit the model
- Step 4: Extract and create a regular gradient of the covariate of interest
- Step 5: Calculate the f parameters as a linear function of covariates from the model

$$
\begin{aligned}
f_1 &= \mathbf{x}'_\alpha \boldsymbol{\alpha} \\
f_2 &= \mathbf{x}'_\beta \boldsymbol{\beta} \\
f_{12} &= \mathbf{x}'_\gamma \boldsymbol{\gamma},
\end{aligned}
$$

Covariate vector

Slope parameters estimated by the model

Manipulating MCMC output in R

- Step 6: Calculate the probabilities of each unique capture history

- Step 7: Calculate ψ: Divide the probability of each capture history by the sum of capture history probabilities

- Step 8: Calculate the marginal or conditional ψ probabilities for each species

- Step 9: Make a dataframe of results

- Step 10: Graph

$$\psi_{11} = \frac{\exp(f_1 + f_2 + f_{12})}{1 + \exp(f_1) + \exp(f_2) + \exp(f_1 + f_2 + f_{12})},$$

$$\psi_{10} = \frac{\exp(f_1)}{1 + \exp(f_1) + \exp(f_2) + \exp(f_1 + f_2 + f_{12})},$$

$$\psi_{01} = \frac{\exp(f_2)}{1 + \exp(f_1) + \exp(f_2) + \exp(f_1 + f_2 + f_{12})},$$

$$\psi_{00} = \frac{1}{1 + \exp(f_1) + \exp(f_2) + \exp(f_1 + f_2 + f_{12})}.$$

Example: How **marginal occupancy** probability varies with housing density

- A model with 3 interacting species: Bobcat, coyote and red fox

- All natural parameters modeled as:

    ~1+Housing_Density5km+Disturbed5km

- f_{123} set to 0

Intro to ggplot2

- A data visualization package in R created by Hadley Wickham

- Plots are constructed in "Layers"

- **Components to a layer are:**
  - **Data:** What you want to graph
  - **Aesthetic mappings:** descriptions of how the variables in the dataset are mapped to visual objects

- You can add many layers to your plot using the "+" sign

- You can also add other controls like colors, backgrounds and change text using a "theme" layer

```
p1 <- ggplot(data = iris, aes(x = Petal.Length, y = Petal.Width))+
    geom_point(aes(color = Species))+
    geom_smooth(method='lm')+
    xlab("Petal Length (cm)") +
    ylab("Petal Width (cm)") +
    ggtitle("Petal Length versus Petal Width")
```

The base plot with data
and aesthetics (defining
the x and y variables in
the data you want to plot.

```
p1 <- ggplot(data = iris, aes(x = Petal.Length, y = Petal.width))+
  geom_point(aes(color = Species))+
  geom_smooth(method='lm')+
  xlab("Petal Length (cm)") +
  ylab("Petal Width (cm)") +
  ggtitle("Petal Length versus Petal Width")
```

A geom layer using the same x and y as the base plot but adding a color aesthetic to represent each species

```r
p1 <- ggplot(data = iris, aes(x = Petal.Length, y = Petal.width))+
   geom_point(aes(color = Species))+
   geom_smooth(method='lm')+
   xlab("Petal Length (cm)") +
   ylab("Petal Width (cm)") +
   ggtitle("Petal Length versus Petal Width")
```

A new layer adding a regression line and confidence interval based on least squares mean

```r
p1 <- ggplot(data = iris, aes(x = Petal.Length, y = Petal.Width))+
  geom_point(aes(color = Species))+
  geom_smooth(method='lm')+
  xlab("Petal Length (cm)") +
  ylab("Petal Width (cm)") +
  ggtitle("Petal Length versus Petal Width")
```

Axis labels and title
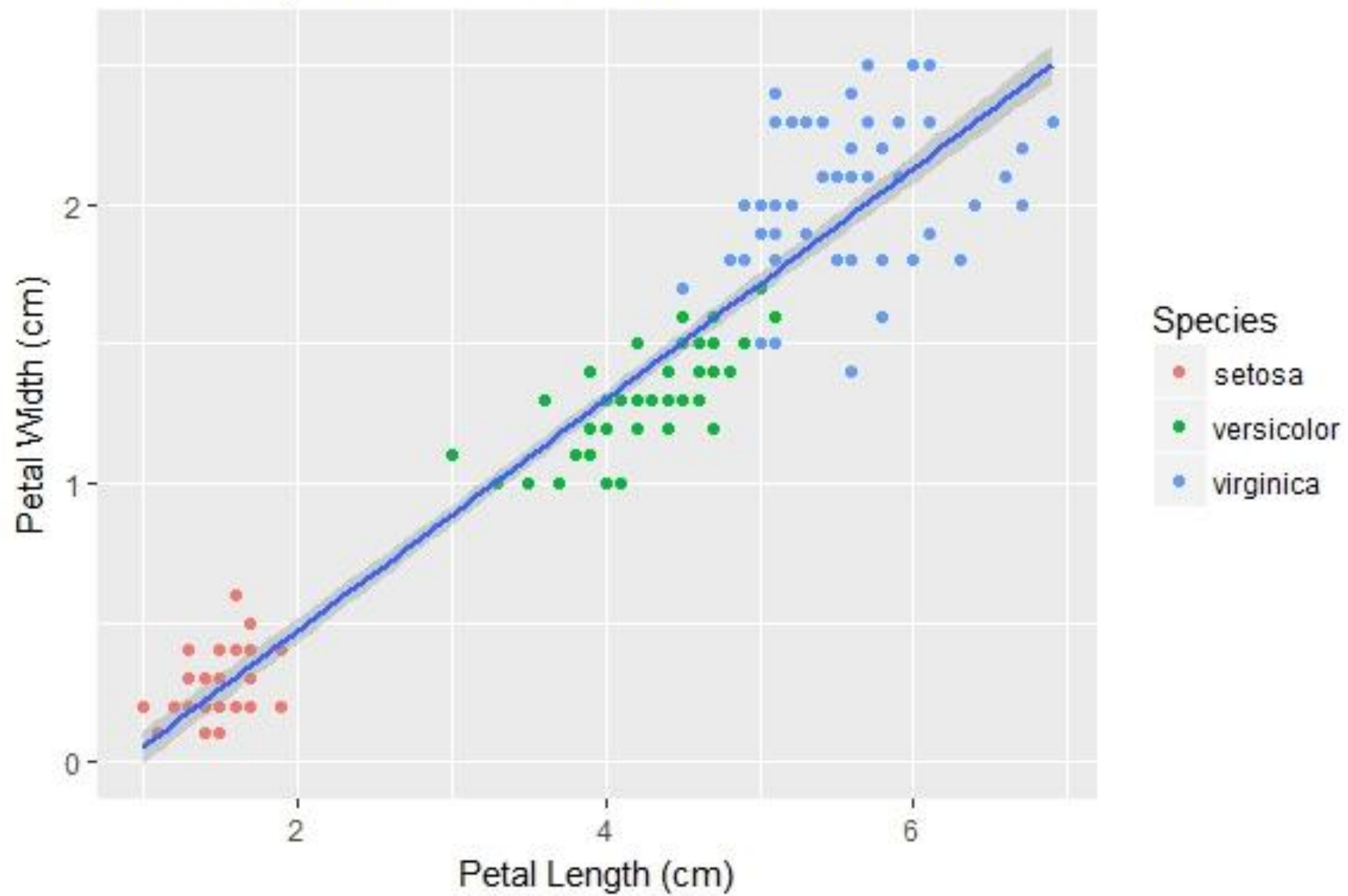
How does marginal occupancy probability vary with Percent Disturbed in a 5km radius?

Hint:

- You will need to now hold housing density at its mean and vary Dist5km when calculating your f parameters
- Probability calculations should not change from the previous example

Try it yourself, we'll go over it together in ~15min

Example: How **conditional occupancy** probability varies with housing density

- Now, we will need to calculate probabilities for each species based on whether or not a given interacting species was present.

Occupancy of Species 1 *in the absence* of Species 2 : $\psi_{10} / \psi_{10} + \psi_{00}$

Occupancy of Species 1 *in the presence* of Species 2: $\psi_{11} / \psi_{11} + \psi_{01}$

How does conditional occupancy probability vary with Percent Disturbed in a 5km radius?

Hint:

- You will need to now hold housing density at its mean and vary Dist5km when calculating your *f* parameters
- Probability calculations should not change from the previous example

Try it yourself, we'll go over it together in ~15min

Summary

- Graphing marginal and conditional probabilities involves calculating the f parameters from a linear combination of covariates and slope parameters from MCMC

- Then, calculation of the probabilities themselves can follow, based on probability rules

- ggplot2 is a flexible R package for graphing