Today

- Ask me anything
  - questions from homework?
  - Rosenbluth algorithm
- Dataset to analyze

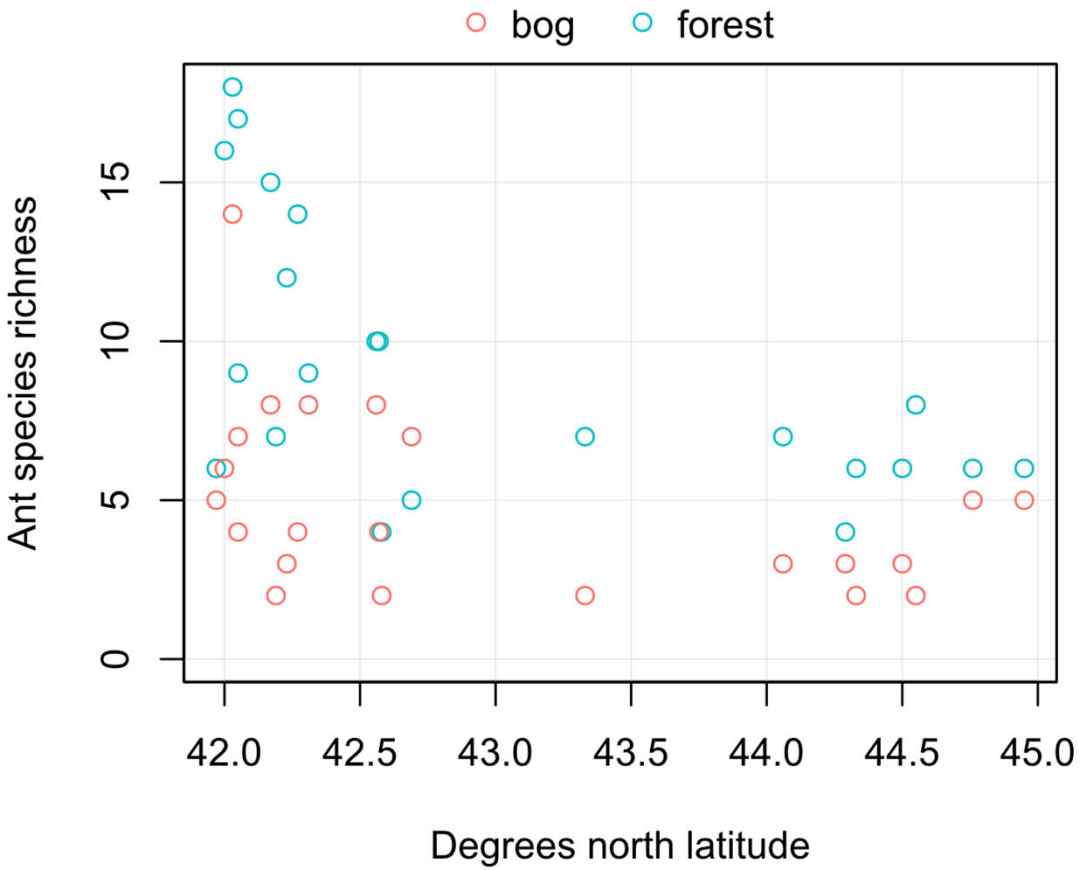
Where are we at?

- Frequentist
  - sampling distribution, SSQ, lm, optim
- Likelihood
  - the likelihood, MLE via optim, likelihood ratios
- Bayesian
  - posterior distribution, prior, MCMC, sampost, ulam, rstanarm
- Simple linear model

Where are we going?

- GLM
  - glm, stan_glm
- GLMM
  - multilevel models, glmer, stan_glmer
- Prediction
  - cross-validation, AIC
  - segue to machine learning

Dataset to analyze



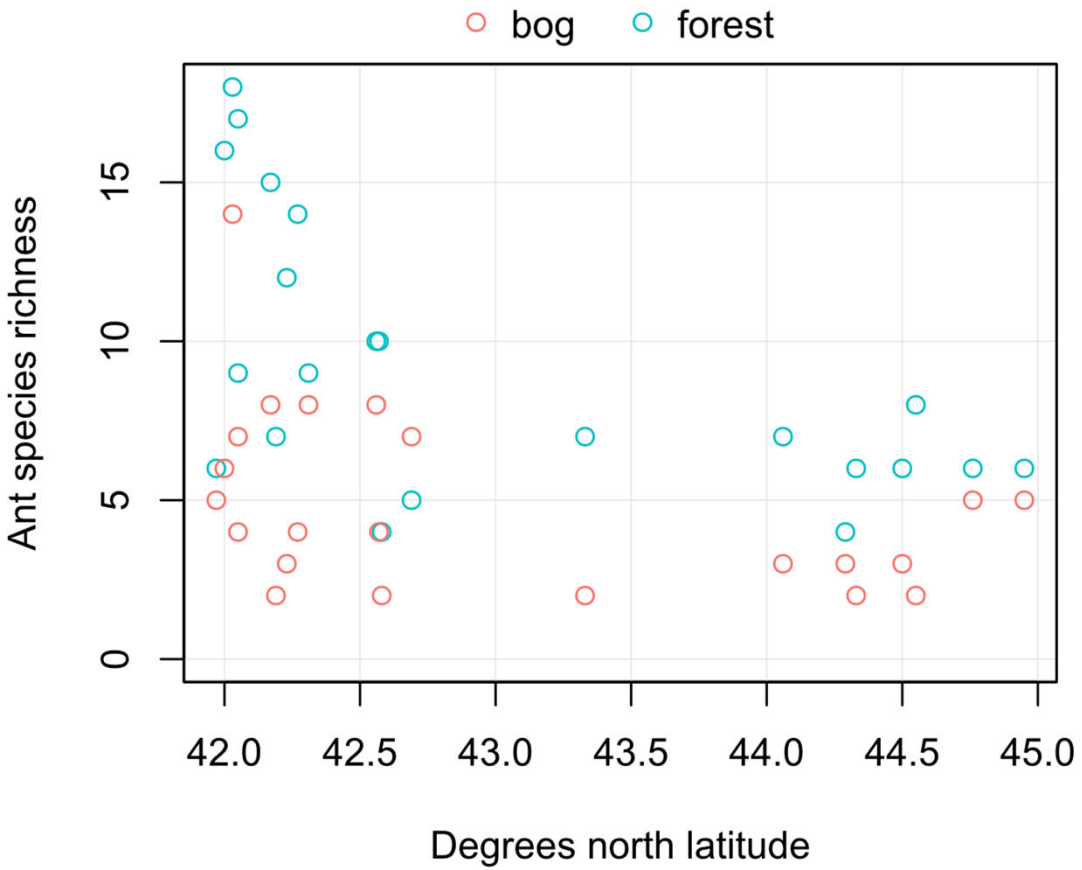
Scientific questions:

How does species richness vary with latitude?

Is this relationship different between habitats?

How different is species richness between habitats?

Dataset to analyze



Write a model

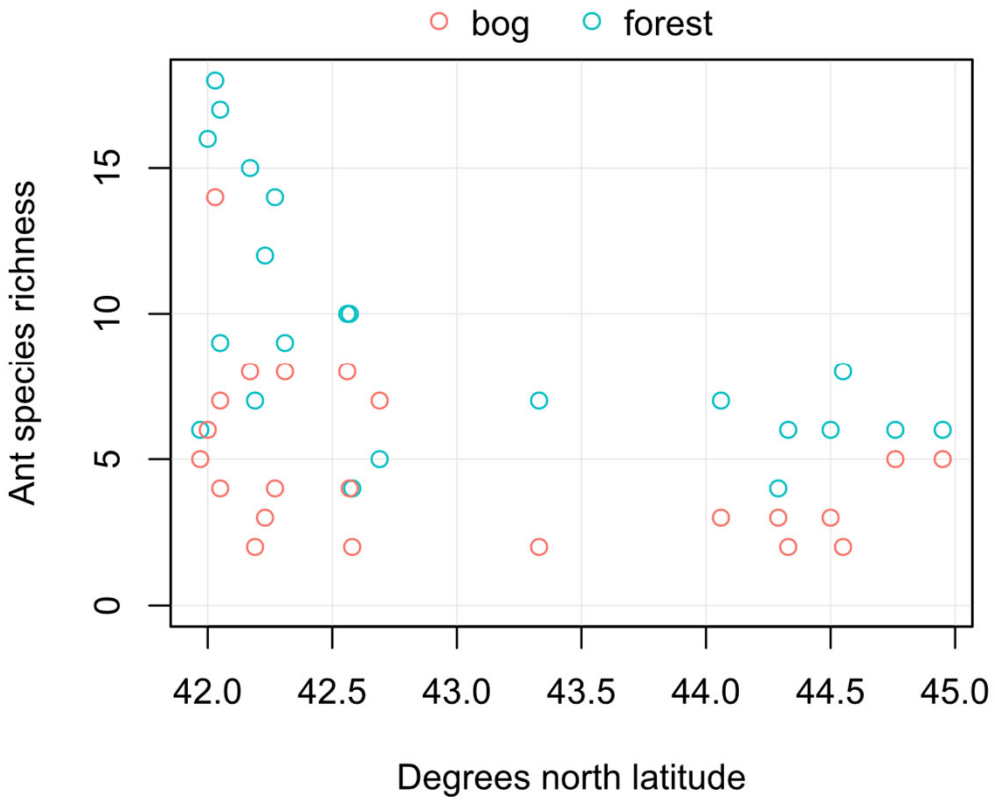
Write a model

$$y_i \sim \text{Normal}(\mu_i, \sigma)$$

$$\mu_i = \begin{cases} \beta_{0,b} + \beta_{1,b} x_i & : \text{bog} \\ \beta_{0,f} + \beta_{1,f} x_i & : \text{forest} \end{cases}$$

y is richness
x is latitude

Assumptions?



$$y_i \sim \text{Normal}(\mu_i, \sigma)$$

$$\mu_i = \begin{cases} \beta_{0,b} + \beta_{1,b} x_i & : \text{bog} \\ \beta_{0,f} + \beta_{1,f} x_i & : \text{forest} \end{cases}$$

Assumptions

- The model (and study design) can answer the research questions
  - how do the scientific questions map to the model?
  - what estimates or predictions of the model answer the research questions?
- The model can serve as a simplification of reality
  - how important are missing details?

Assumptions

Normal linear model

1. Linearity! (of richness vs latitude)
2. y (richness) is continuous $-\infty$ to $+\infty$
   (richness can be negative)
3. x (latitude) is measured without error
4. Normality of y (richness) given $\mu_i = \beta_0 + \beta_1 x_i$
   (or normality of errors in $y_i = \beta_0 + \beta_1 x_i + e_i$)
5. Errors are identically distributed
   (i.e. same σ, AKA homoscedastic)
6. Independence of errors (e.g. in space and time)

} iid