

# Today

- Miscellaneous
  - portfolio checklist
  - Marktext application (withdraw my recommendation; silent edits)
- Diagnostics
  - ants example

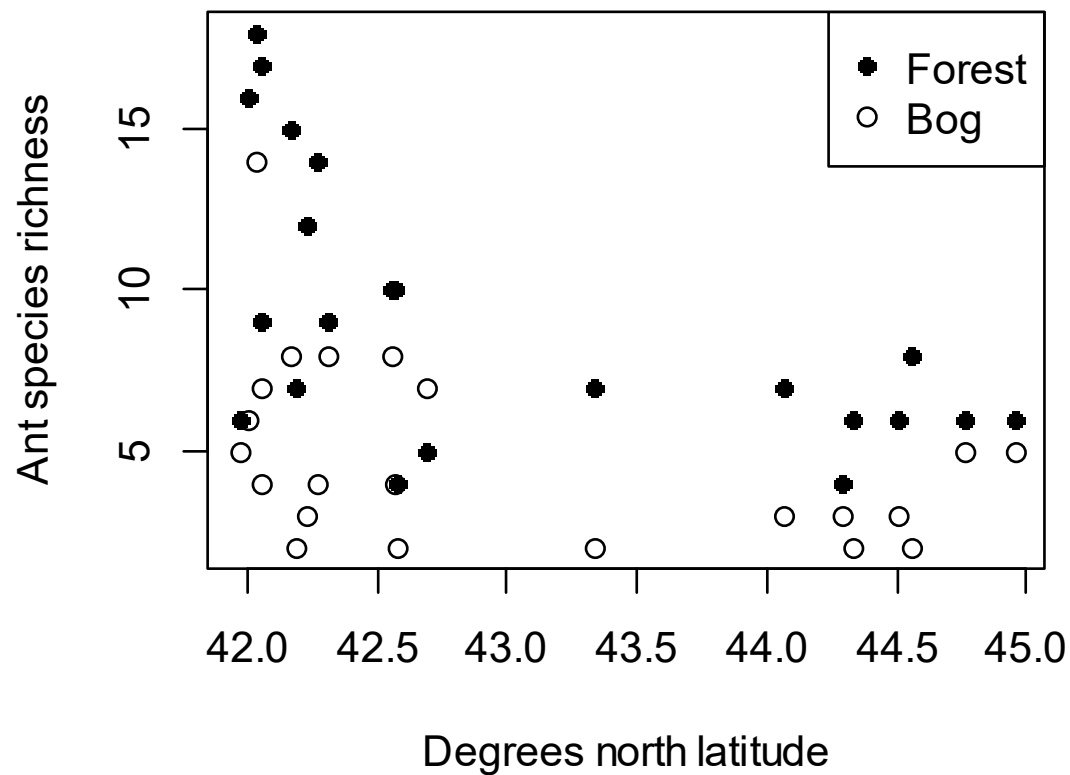
# Diagnostics: model checking

- Systematic departures of the process (biological) model from the data
- Poor error distribution
- Mistakes in data
- Outliers
- Influential data points

# Tools

- Plot the fitted model with the data
- Residuals vs fitted values
- QQ plot, histogram of residuals
- Leave one out (LOO) influence algorithm
- What should these diagnostics look like (on average and variation)?
  - make plots of them from simulated data of the fitted model

# Ants: diagnostics



Write model  
Assumptions?

# Write the model

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

$$\mu_i = \beta_{0,b} + \beta_{1,b}x_i \quad : \text{bog}$$

$$\beta_{0,f} + \beta_{1,f}x_i \quad : \text{forest}$$

y is richness

x is latitude

# 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  $\sigma$ , AKA homoscedastic)
6. Independence of errors (e.g. in space and time)

iid

# Factor

```
ant$habitat <- factor(ant$habitat)
```

```
print(ant$habitat)
```

```
[1] forest forest forest forest forest ...  
[14] forest forest forest forest forest ...  
[27] bog      bog      bog      bog      bog      bog ...  
[40] bog      bog      bog      bog      bog  
Levels: bog forest
```

R data structure for categorical variables; ?factor  
Attribute: levels (sorted alphabetically by default)