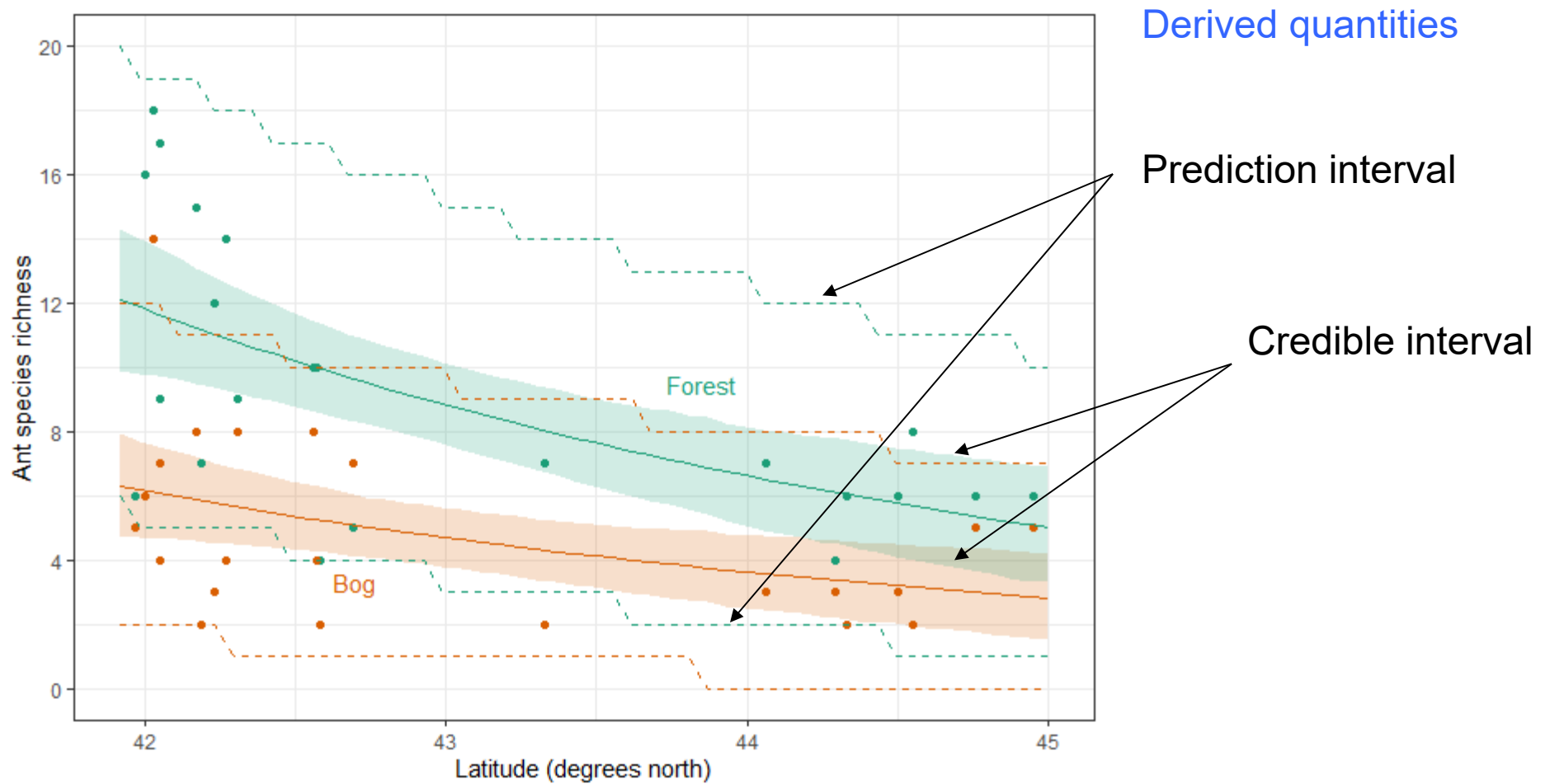


Today

- Ants GLM, Bayesian (rethinking)
 - working with posterior samples
 - derived quantities
 - Answer to: How different is species richness between habitats?
- Ants GLM, frequentist
 - model formulae
- Ants GLM, Bayesian (rstanarm)
 - model checking

Bayesian model - ants



```
# Initialize a grid of latitudes, scaled the same as we scaled the data
lat_upr <- (45 - mean_lat) / sd_lat
lat_lwr <- (41.92 - mean_lat) / sd_lat
latitude <- seq(from=lat_lwr, to=lat_upr, length.out=50)
```

Keep scaling factors

```
# Initialize storage
n <- length(latitude)
hpdi_bog <- matrix(NA, nrow=n, ncol=5) #to store hpdi values and mean
colnames(hpdi_bog) <- c("mnmu", "mulo95", "muhi95", "ppdlo95", "ppdhi95")
hpdi_forest <- matrix(NA, nrow=n, ncol=5)
colnames(hpdi_forest) <- c("mnmu", "mulo95", "muhi95", "ppdlo95", "ppdhi95")
```

```
# For each latitude, form the posterior
for ( i in 1:n ) {
```

```
  # First form samples for the linear predictor \eta
```

```
  eta_bog <- samples$beta_0 +
    samples$beta_2 * latitude[i]
  eta_forest <- samples$beta_0 +
    samples$beta_1 +
    samples$beta_2 * latitude[i] +
    samples$beta_3 * latitude[i]
```

Model on linear predictor scale

Sampled parameters (vectors 36000)

```
  # Then use inverse link for samples of the posterior \mu
```

```
  mu_bog <- exp(eta_bog)
  mu_forest <- exp(eta_forest)
```

Mean species richness

```
  # Sample from Poisson to get the posterior predictive distribution
```

```
  ppd_bog <- rpois(n=length(mu_bog), lambda=mu_bog)
  ppd_forest <- rpois(n=length(mu_forest), lambda=mu_forest)
```

Simulate data generating process
Once for each sample

```
  # Mean and intervals of these samples
```

```
  hpdi_bog[i,1] <- mean(mu_bog)
  hpdi_bog[i,2:3] <- HPDI(mu_bog, prob=0.95)
  #hpdi_bog[i,4:5] <- HPDI(ppd_bog, prob=0.95)
  hpdi_bog[i,4:5] <- quantile(ppd_bog, prob=c(0.025,0.975)) #CPI
  hpdi_forest[i,1] <- mean(mu_forest)
  hpdi_forest[i,2:3] <- HPDI(mu_forest, prob=0.95)
  #hpdi_forest[i,4:5] <- HPDI(ppd_forest, prob=0.95)
  hpdi_forest[i,4:5] <- quantile(ppd_forest, prob=c(0.025,0.975)) #CPI
```

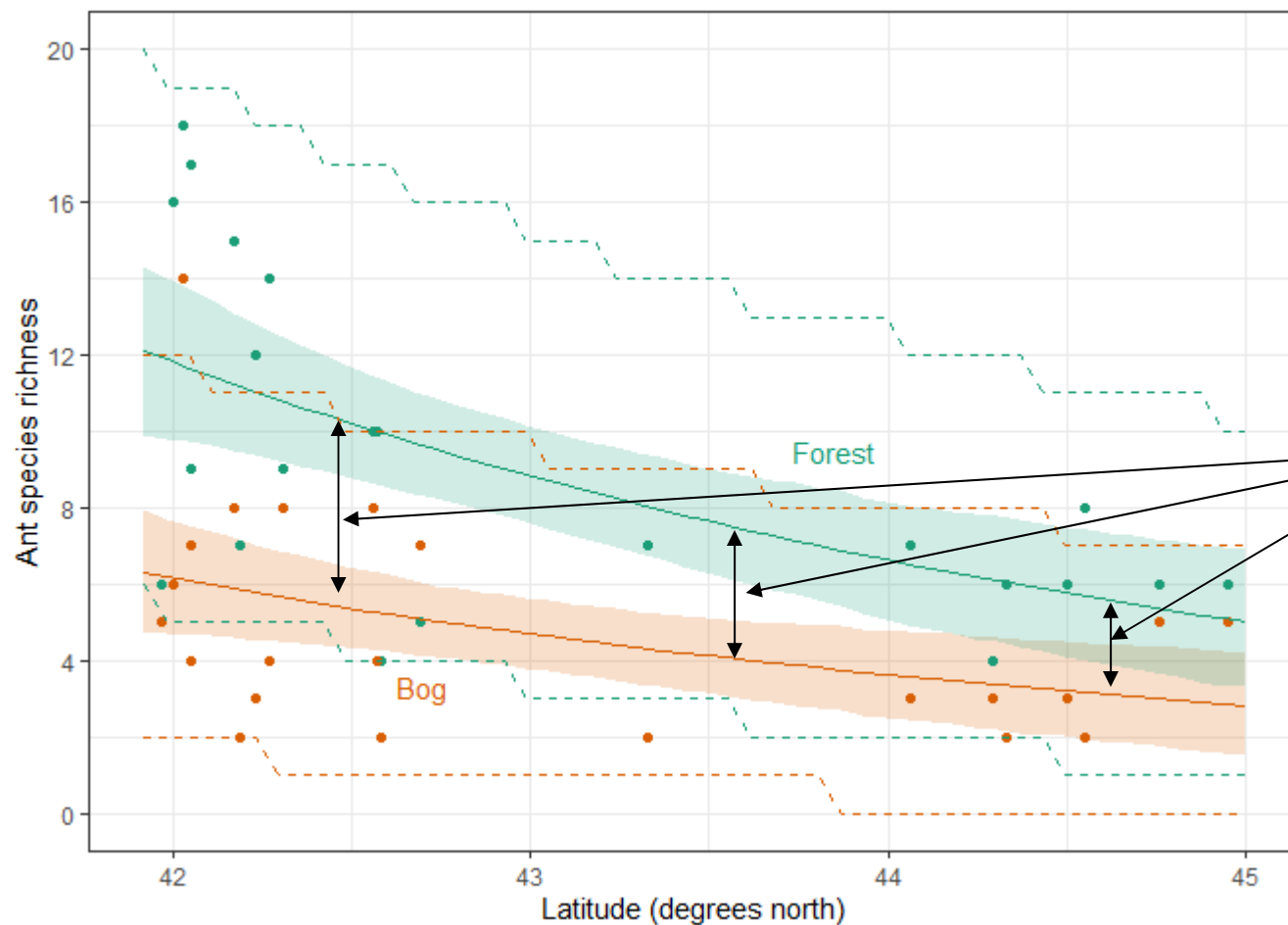
```
}
```

Package in tidy format for plotting

```
latitude <- latitude * sd_lat + mean_lat  
predsbog <- data.frame(habitat=rep("bog", n), latitude, hpdi_bog)  
predsforest <- data.frame(habitat=rep("forest", n), latitude, hpdi_forest)  
preds <- rbind(predsbog, predsforest)
```

Reverse the standardization

Bayesian model - ants



How different is
species richness
between habitats?

Derived quantity

Differences at
different latitudes

```

# Initialize variables and storage
lat_lwr <- (41.92 - mean_lat) / sd_lat
lat_upr <- (45 - mean_lat) / sd_lat
latitude <- seq(from=lat_lwr, to=lat_upr, length.out=50)
n <- length(latitude)
forest_bog_diff <- matrix(NA, nrow=n, ncol=3) #to store mean and hpdi values
colnames(forest_bog_diff) <- c("mndiff","difflo95","diffhi95")

# For each latitude, form the posterior
for ( i in 1:n ) {

  # First form samples for the linear predictor \eta
  eta_bog <- samples$beta_0 +
    samples$beta_2 * latitude[i]
  eta_forest <- samples$beta_0 +
    samples$beta_1 +
    samples$beta_2 * latitude[i] +
    samples$beta_3 * latitude[i]

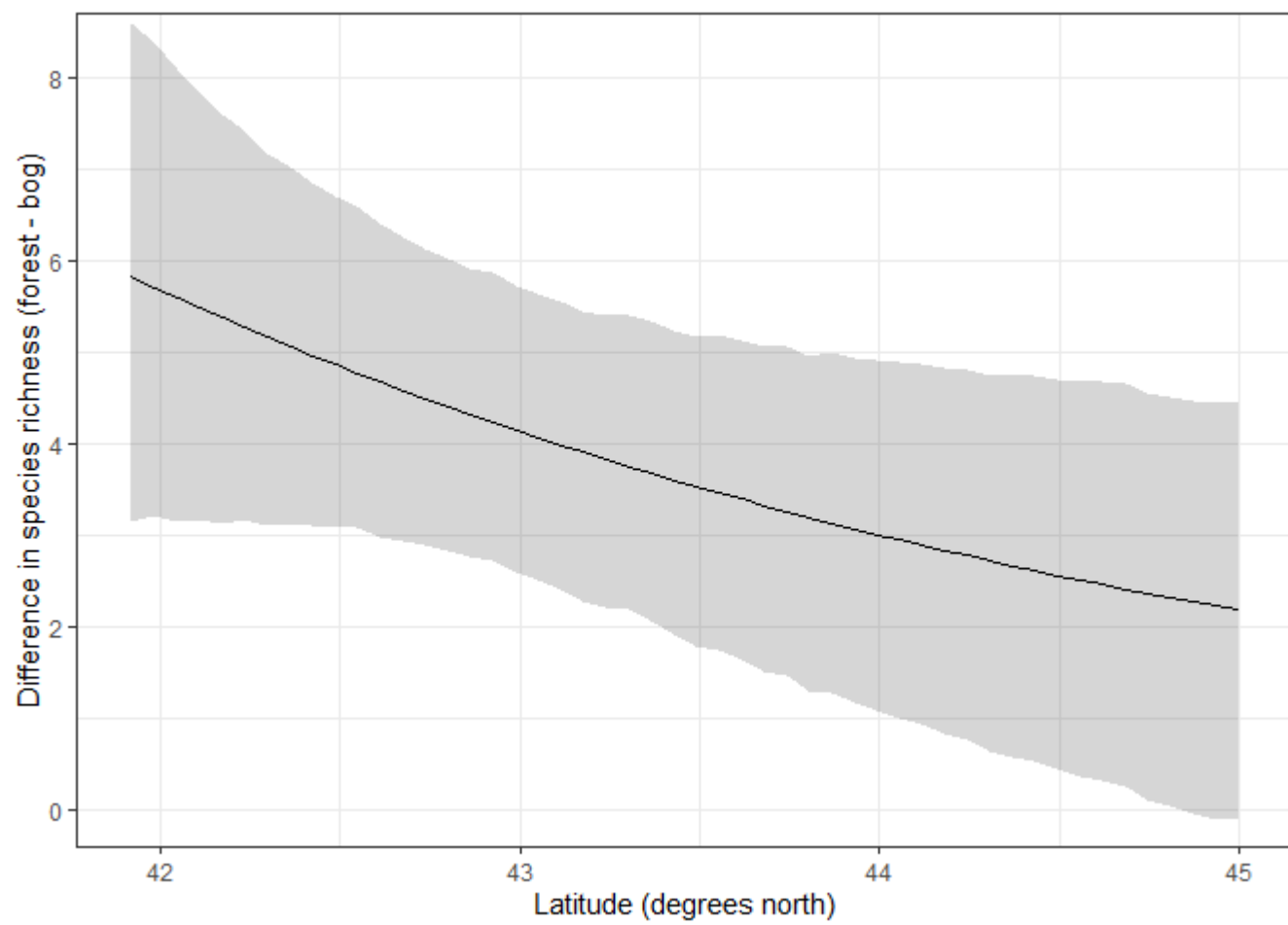
  # Then use inverse link for samples of the posterior \mu
  mu_bog <- exp(eta_bog)
  mu_forest <- exp(eta_forest)

  # Now calculate the habitat difference (derived quantity)
  diff <- mu_forest - mu_bog

  # Mean and intervals of these samples
  forest_bog_diff[i,1] <- mean(diff)
  forest_bog_diff[i,2:3] <- HPDI(diff, prob=0.95)
  #forest_bog_diff[i,2:3] <- quantile(diff, prob=c(0.025,0.975)) #CPI
}

```

Here's the derived quantity. Output: 36000 samples.



Model formulae

$$y \sim x1 * x2 + x3 / x4$$

equivalent to:

$$y \sim 1 + x1 + x2 + x1:x2 + x3 + x3:x4$$

Model formula

```
fit <- glm(richness ~ habitat + latitude + habitat:latitude,  
          family=poisson(link="log"), data=ant)
```

$$\eta_i = \beta_0 \text{intercept}_i + \beta_1 \text{forest}_i + \beta_2 \text{latitude}_i + \beta_3 \text{forest}_i \times \text{latitude}_i$$

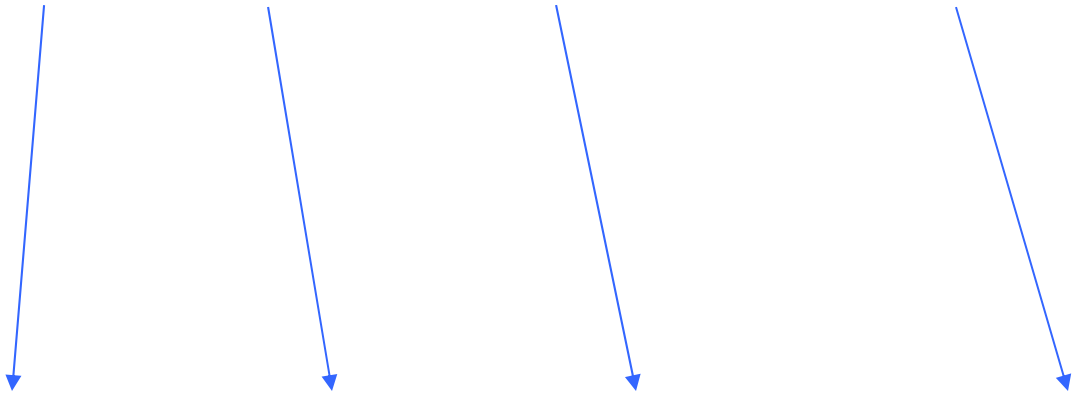
Model formula

```
fit <- glm(richness ~ habitat + latitude + habitat:latitude,  
          family=poisson(link="log"), data=ant)
```

Equivalent:

```
richness ~ habitat * latitude
```

```
richness ~ 1 + habitat + latitude + habitat:latitude
```


$$\eta_i = \beta_0 intercept_i + \beta_1 forest_i + \beta_2 latitude_i + \beta_3 forest_i \times latitude_i$$

Model matrix (design matrix)

```
fit <- glm(richness ~ habitat + latitude + habitat:latitude,
           family=poisson(link="log"), data=ant)
```

Data

habitat	latitude	richness
forest	42	16
forest	42.56	10
forest	43.33	7
forest	44.76	6
bog	42.17	8
bog	42.57	4
bog	44.06	3
bog	44.95	5

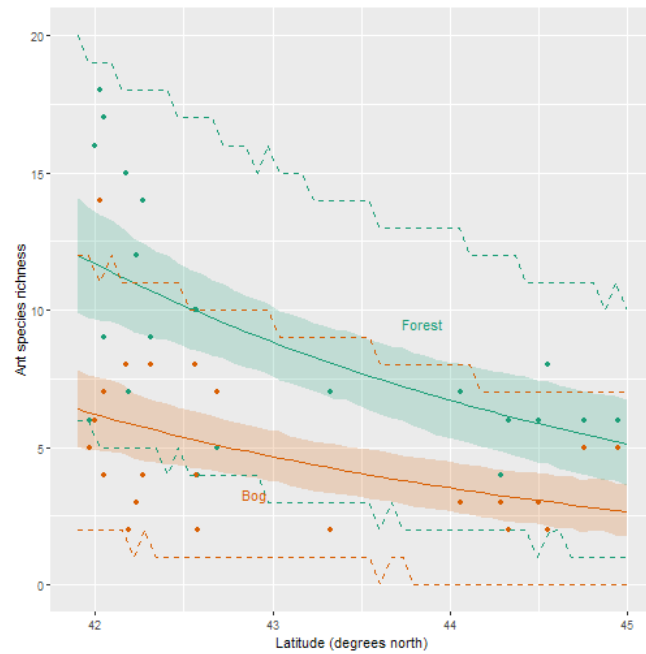
Model matrix

`model.matrix(fit)`

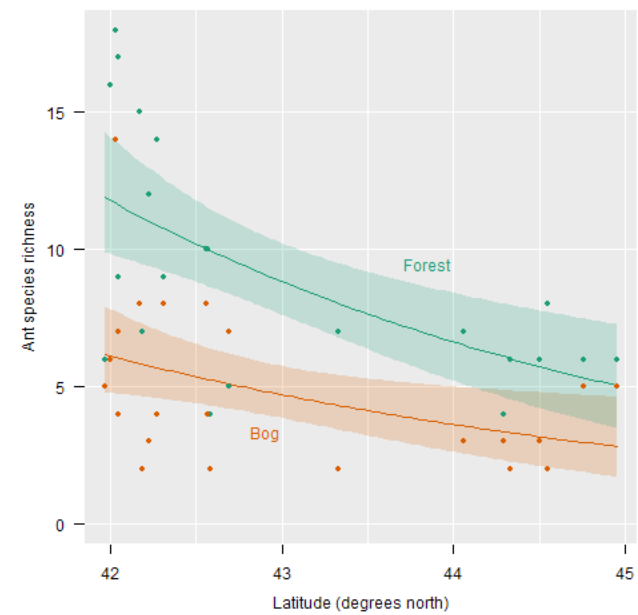
intercept	forest	latitude	forest:latitude
1	1	42	42
1	1	42.56	42.56
1	1	43.33	43.33
1	1	44.76	44.76
1	0	42.17	0
1	0	42.57	0
1	0	44.06	0
1	0	44.95	0

$$\eta_i = \beta_0 \text{intercept}_i + \beta_1 \text{forest}_i + \beta_2 \text{latitude}_i + \beta_3 \text{forest}_i \times \text{latitude}_i$$

eta ~ 1 + forest + latitude + forest:latitude



Bayesian



Frequentist
(w approx intervals)

Comparing inference algorithms for frequentist and Bayesian approaches to model means and predictions so far:

Tool	Mean	Uncertainty of mean	Uncertainty of prediction
lm	predict()	predict(int="confidence")	predict(int="prediction")
glm	predict(type= "response")	predict(se.fit=TRUE)	via bootstrap
		or via bootstrap	
stan_glm	mean(pmu)	hpdi(pmu), cpi(pmu)	hpdi(ppd), cpi(ppd)

where:

- `pmu <- posterior_linpred(transform = TRUE)` , OR `pmu <- posterior_epred()`
- `ppd <- posterior_predict()`