

Tidy data

country	year	cases	population
Afghanistan	1999	745	19987071
Afghanistan	2000	2666	20095360
Brazil	1999	37737	172006362
Brazil	2000	80488	174004898
China	1999	210258	1272015272
China	2000	210766	1280423583

**Variables
in
columns**

country	year	cases	population
Afghanistan	1999	745	19987071
Afghanistan	2000	2666	20095360
Brazil	1999	37737	172006362
Brazil	2000	80488	174004898
China	1999	210258	1272015272
China	2000	210766	1280423583

**Observations
in
rows**

country	year	cases	population
Afghanistan	99	745	19987071
Afghanistan	00	2666	20095360
Brazil	99	37737	172006362
Brazil	00	80488	174004898
China	99	210258	1272015272
China	00	210766	1280423583

**Values
in
cells**

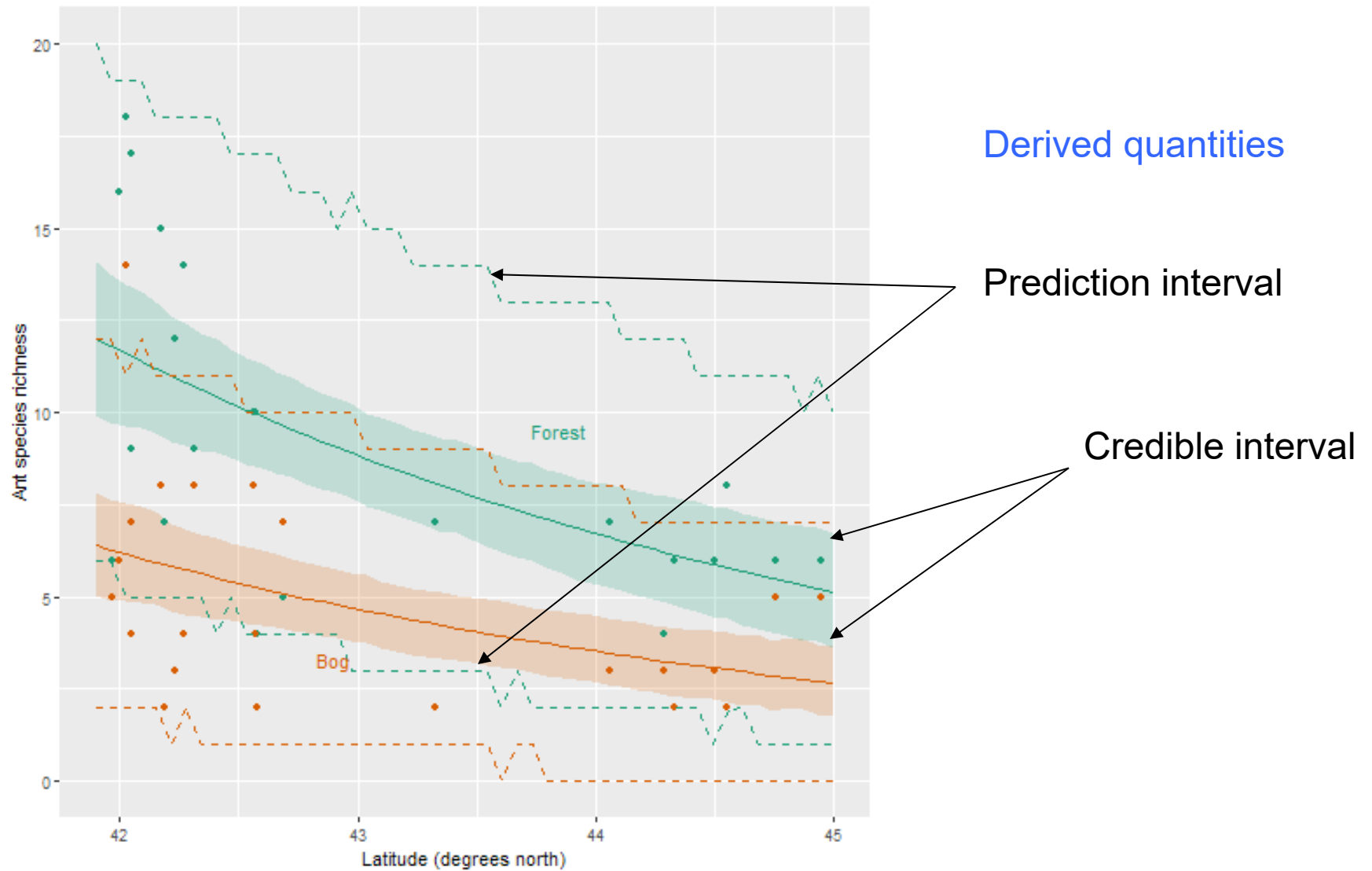
- `pivot_longer()` - tidy a variable that is in multiple columns
- `pivot_wider()` - tidy an observation that is in multiple rows
- `separate()` - split a column into multiple variables
- `unite()` - collapse multiple columns into a single variable

Base R: `reshape()`, `stack()`, `unstack()`, `strsplit()`, `paste()`

Finish off ants GLM

- Bayesian
- Working with posterior samples
- Derived quantities
- Answer to: How different is species richness between habitats?

Bayesian model - ants



```

# Initialize variables and storage
latitude <- seq(from=41.92, to=45, length.out=50) #range for latitude
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$alpha[,1] +
    samples$beta[,2] * latitude[i]
  eta_forest <- samples$alpha[,1] +
    samples$beta[,1] +
    samples$beta[,2] * latitude[i] +
    samples$beta[,3] * latitude[i]

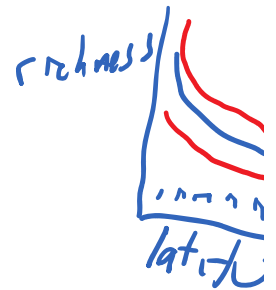
  # Then use inverse link for samples of the posterior \mu
  mu_bog <- exp(eta_bog)
  mu_forest <- exp(eta_forest)

  # Sample from Poisson to get the posterior predictive distribution
  ppd_bog <- rpois(n=length(mu_bog), lambda=mu_bog) - samples
  ppd_forest <- rpois(n=length(mu_forest), lambda=mu_forest)

  # 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
}

```

} setting up storage



$$\eta_{bog} = \alpha + \beta_2 \text{latitude}$$

$$\eta_{forest} = \alpha + \beta_1 + (\beta_2 + \beta_3) \text{latitude}$$

$$\mu_{bog} = e^{\eta_{bog}} \quad \text{link function}$$

$$\mu_{forest} = e^{\eta_{forest}} \quad \log(\mu) = \eta$$

Now have 2 additional columns:
 samples for 1) posterior pred bog
 2) " " " forest

← mean relationship.

← credible for mean relationship.

← posterior predictive interval

```
# Initialize variables and storage
latitude <- seq(from=41.92, to=45, length.out=50) #range for latitude
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")
```

} setup storage.

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

```
  # First form samples for the linear predictor \eta
  eta_bog <- samples$alpha[,1] +
             samples$beta[,2] * latitude[i]
  eta_forest <- samples$alpha[,1] +
                samples$beta[,1] +
                samples$beta[,2] * latitude[i] +
                samples$beta[,3] * latitude[i]
```

} All the same as previous.

```
  # 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
```

← new derived quantity

→ new column of posterior samples

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

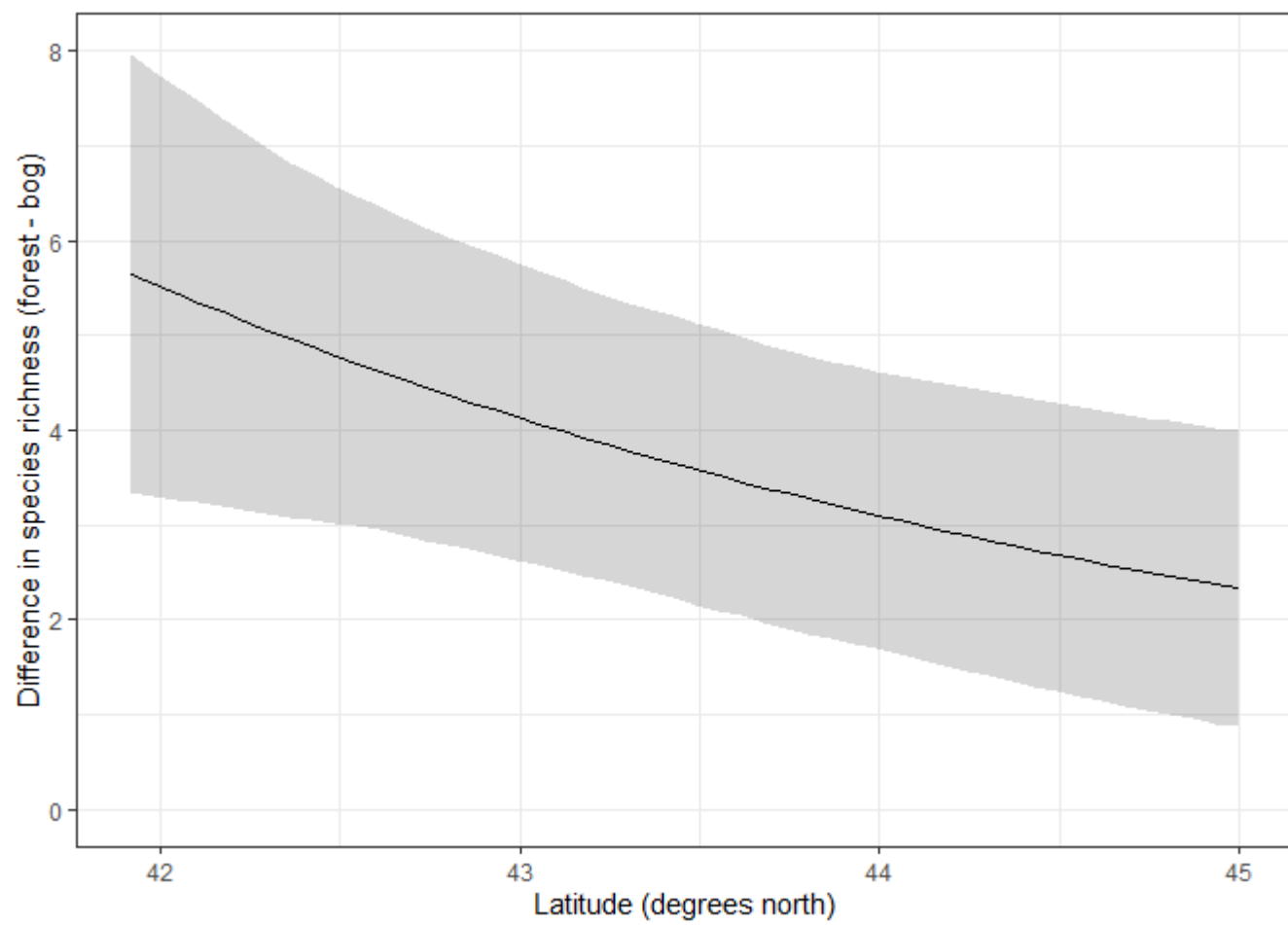
```
  forest_bog_diff[i,1] <- mean(diff)
```

```
  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
```

} credible interval

```
}
```



**We considered EDA for the radon example.
See 10_3_radon_multilevel.md**