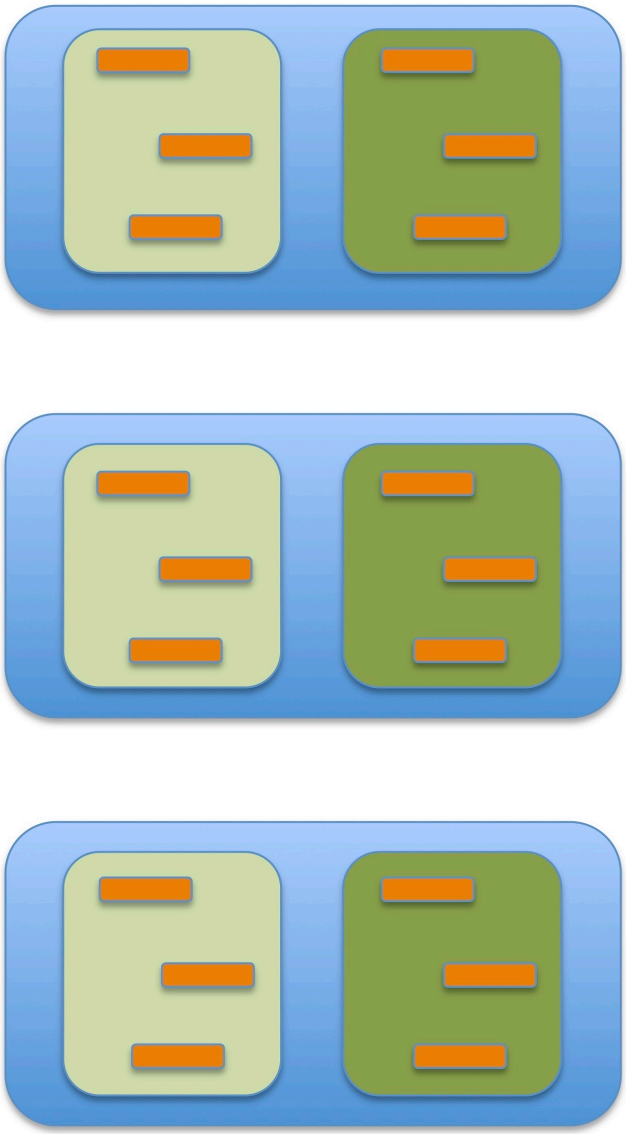
Grassland grazing experiment: repeated measures, nested and crossed random effects, fixed vs. random effects

Jerrentrup et al. 2014: how does low vs moderate cattle grazing affect insect communities?

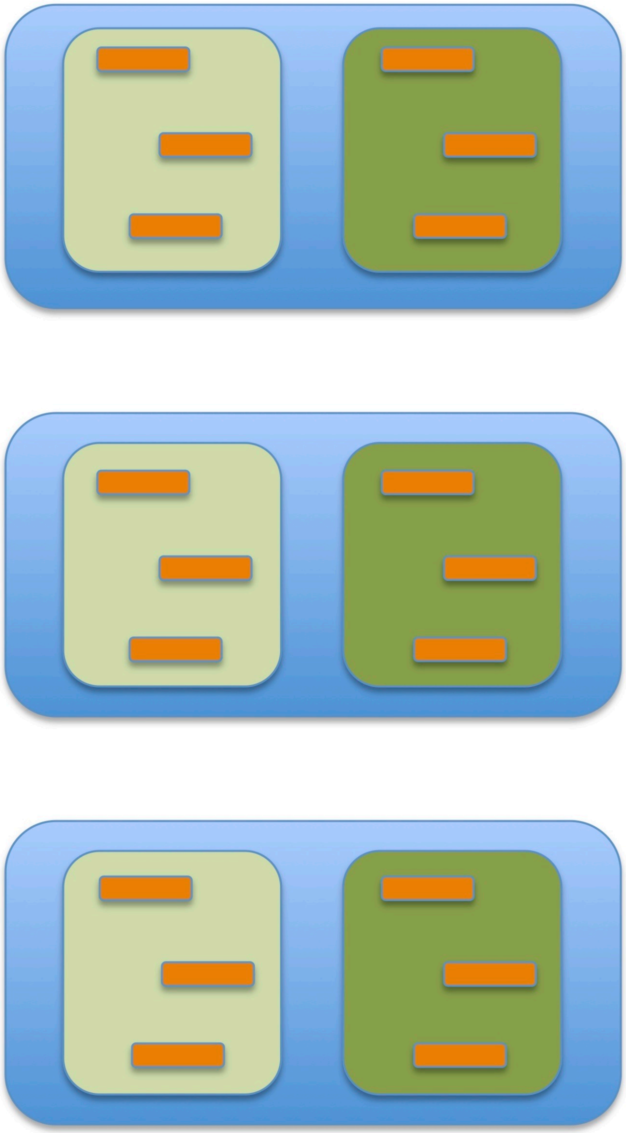


Block

Paddock = Treatment

Transect

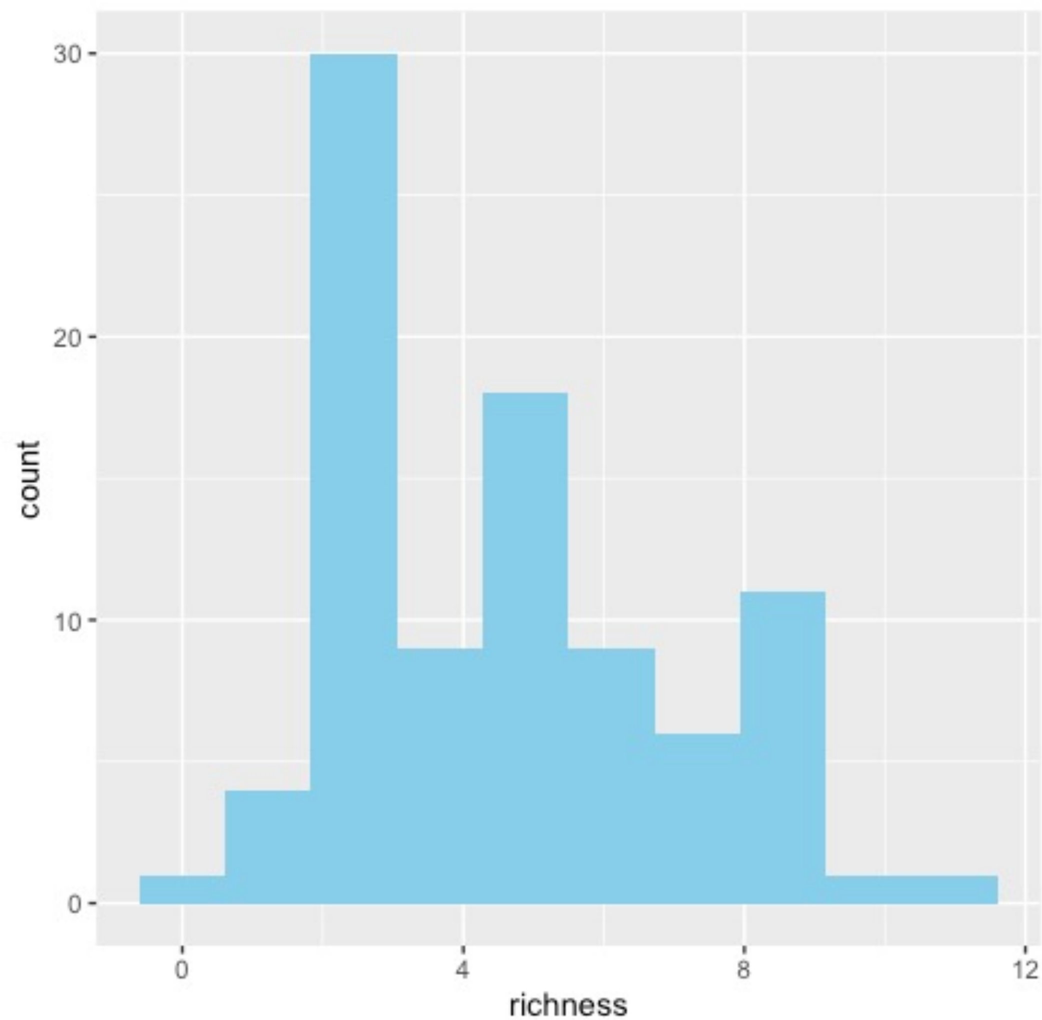
- Butterfly species richness
- 2002, 2003, 2004, 2010, 2011



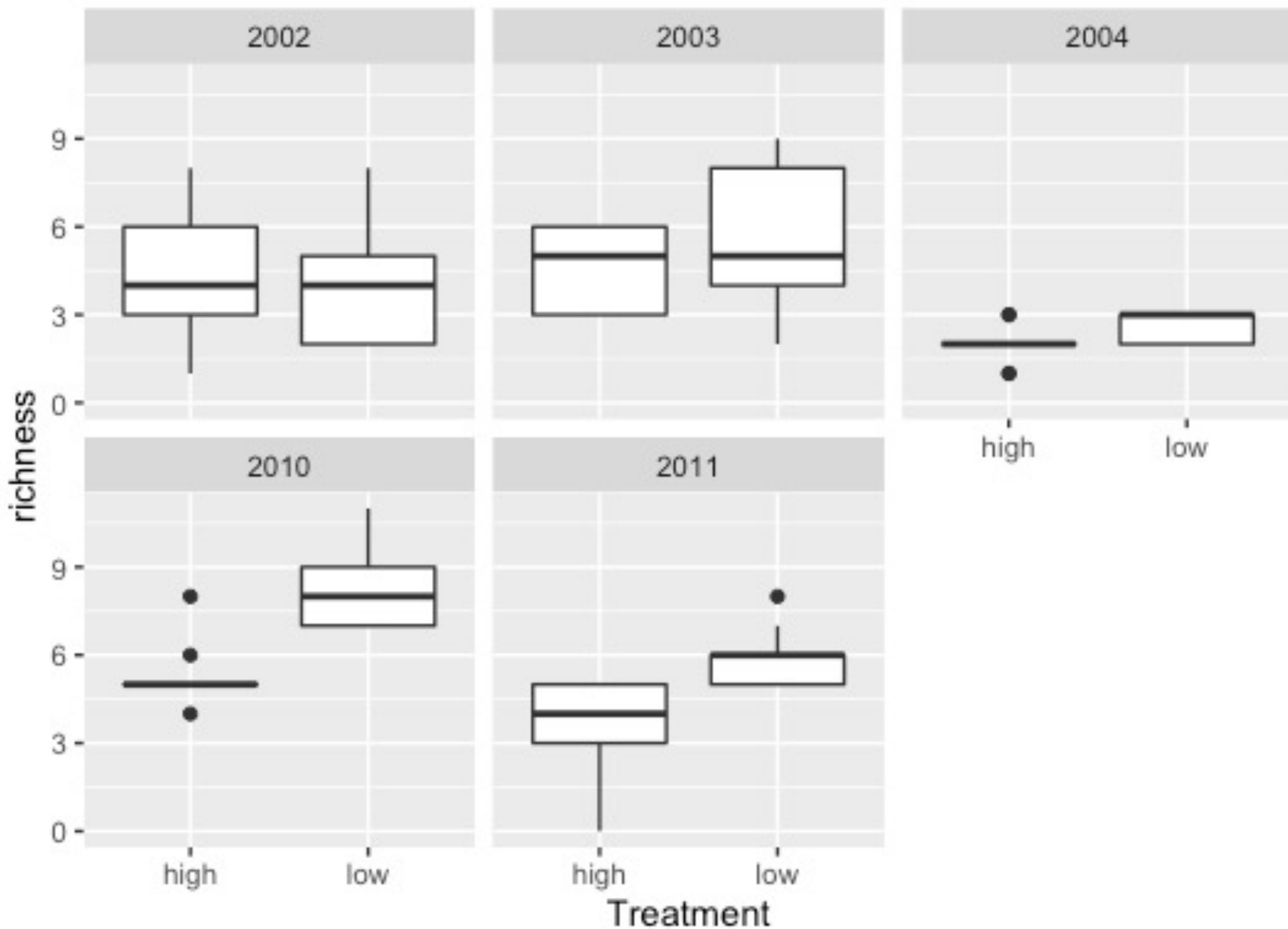
Block

Paddock = Treatment

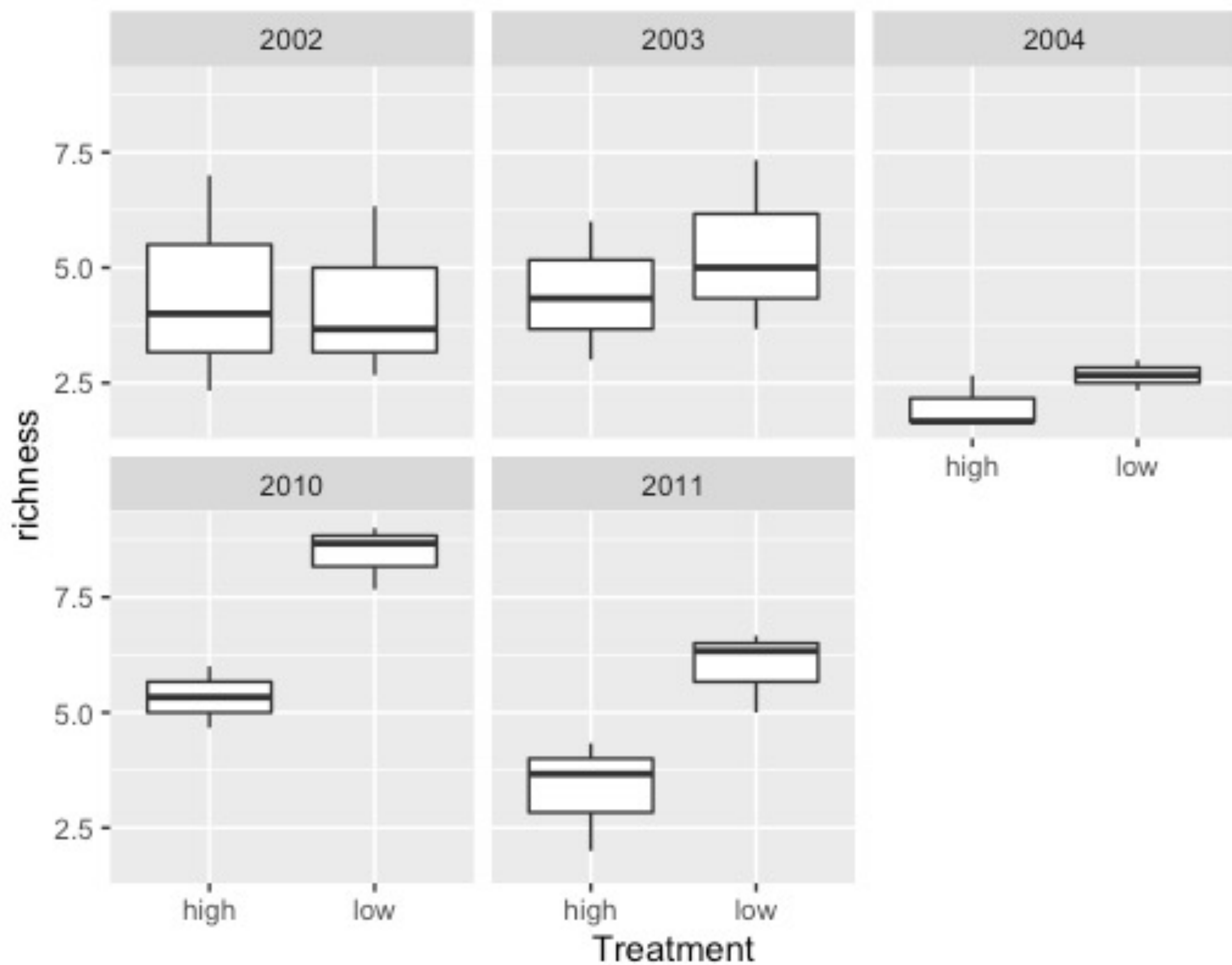
Transect



```
ggplot(butterfly, aes(Treatment, richness)) + geom_boxplot() +
facet_wrap(~Year)
```



```
butterfly.means = ddply(butterfly, .(Year, Block, Treatment,
Paddock), summarize, richness = mean(richness))

ggplot(butterfly.means, aes(Treatment, richness)) + geom_boxplot()
+ facet_wrap(~Year)
```



Nested groups

Blocks, Paddocks, Transects – non-independence + degrees of freedom

Nested random effects: handled automatically, if you give each group its own level

```
butterfly$Transect

##  [1] 1 2 3 1 2 3 1 2 3 1 2 3…

butterfly$TransectID

##  [1] A1_1 A1_2 A1_3 A2_1 A2_2 A2_3 B1_1 B1_2 B1_3 B2_1 B2_2 B2_3… ## 18 Lev
els: A1_1 A1_2 A1_3 A2_1 A2_2 A2_3 B1_1 B1_2 B1_3 B2_1 B2_2 ... C2_3
```

```
(1|Block/Paddock/Transect)
```

```
(1|Block) + (1|Paddock) + (1|TransectID)
```

Nested groups

Do we actually need to include transect-level data?

Or could we just average them as subsamples?

In this case don’t really need this resolution, but will include anyways as an example

If you had transect-level predictors, would be more important

Repeated measures

- Each transect is measured 5 times
- Also true at the level of paddocks and blocks
- Random effects can account for repeated measures, within limits
- (1|TransectID) will account for the fact that observations within a transect will be correlated, in the sense that they are similar to each other
- This does not account for **temporal autocorrelation**: observations close together in time may be more similar
- Most problematic for long time series. For 5 observations, little temporal structure

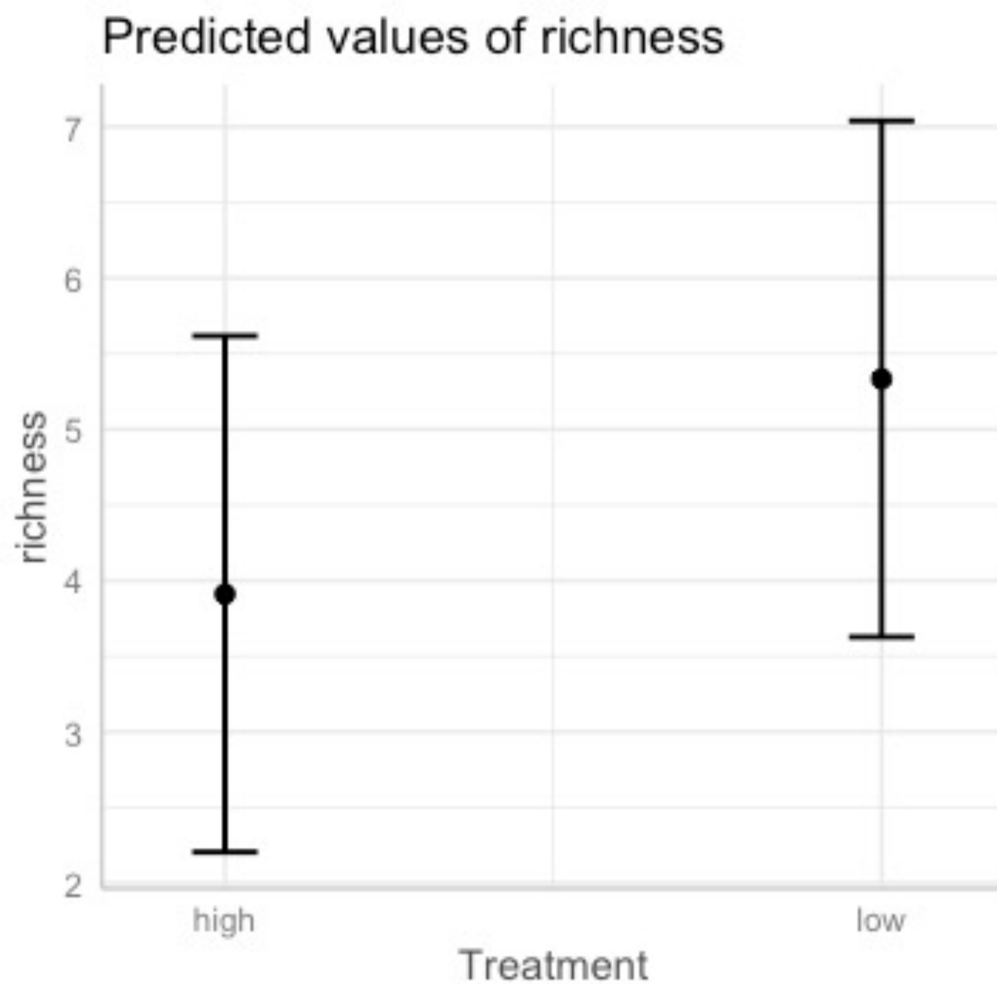
Crossed random effects

- The spatial groups are nested
- But observations from the same year are likely to be similar to each other
- So we need to account for Year as well
- Year will be ‘crossed’ with the spatial effects
- Year effect = difference between years, averaged over space
- Block effect = difference between blocks, averaged over time

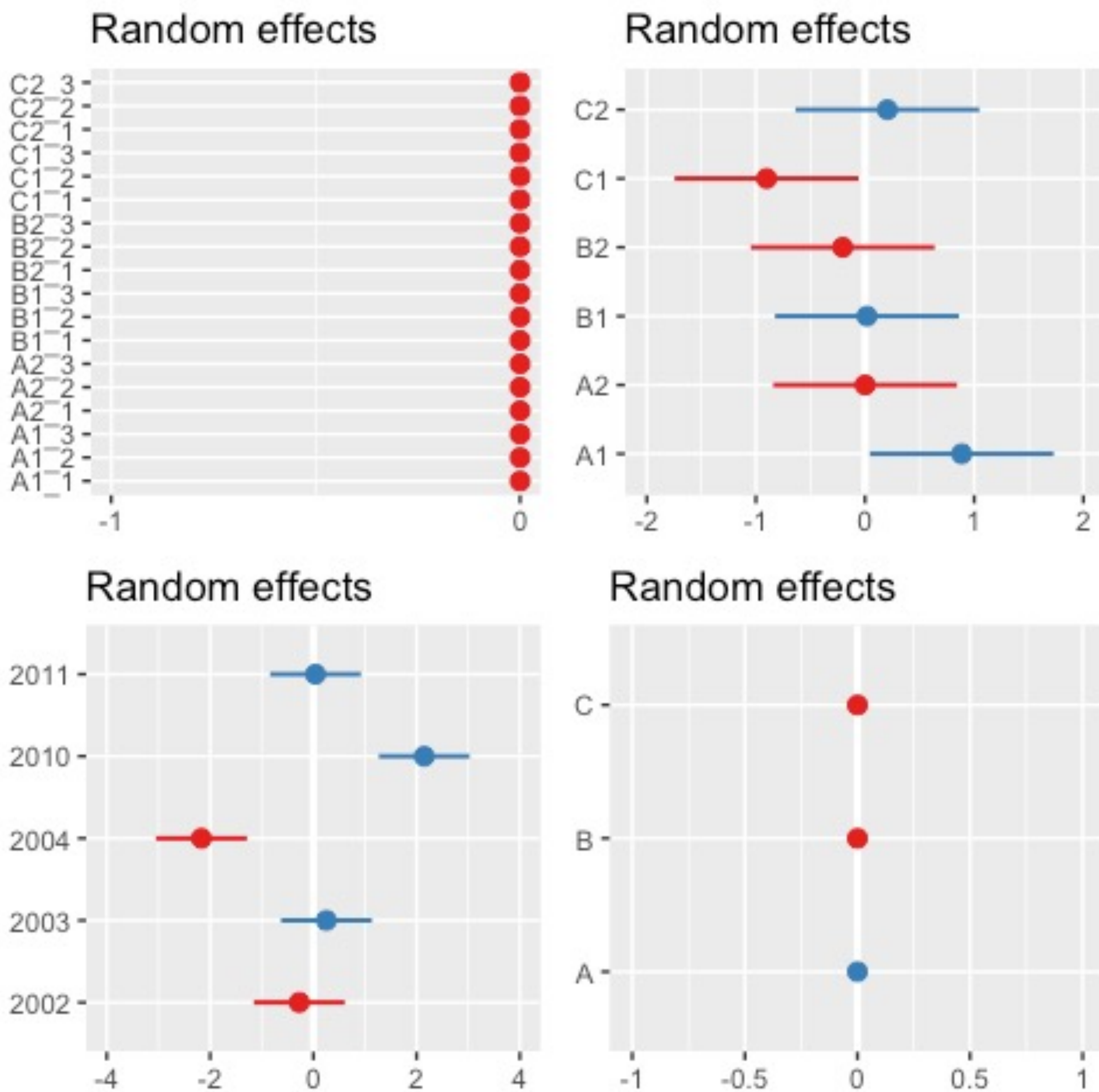
```r
mod.rand = lmer(richness ~ Treatment + (1|Year) + (1|TransectID) + (1|Block) +
(1|Paddock), data = butterfly)
summary(mod.rand)
##
## Random effects:
##  Groups     Name        Variance Std.Dev.
##  TransectID (Intercept) 0.000    0.000
##  Paddock    (Intercept) 0.549    0.741
##  Year       (Intercept) 2.494    1.579
##  Block      (Intercept) 0.000    0.000
##  Residual               2.513    1.585
## Number of obs: 90, groups:  TransectID, 18; Paddock, 6; Year, 5; Block, 3
##
## Fixed effects:
##             Estimate Std. Error    df t value Pr(>|t|)
## (Intercept)    3.911      0.859 6.500    4.55   0.0032 **
## Treatmentlow   1.422      0.691 4.000    2.06   0.1087
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

Zeros for random effects: the variation at this level is not greater than we would expect based on the variation at lower (residual) or higher levels

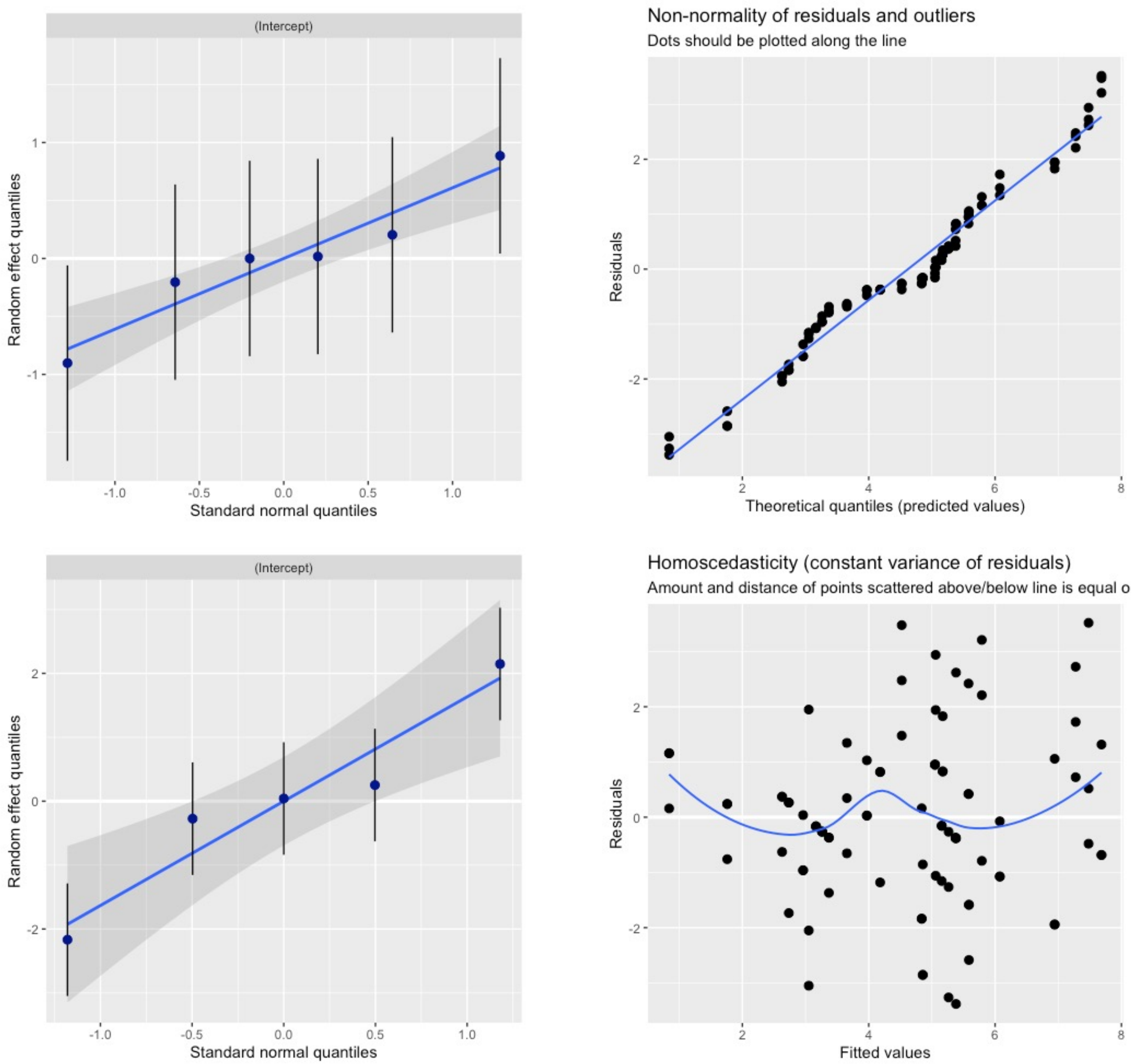
```
plot(ggeffect(mod.rand))
```



```
grid.arrange(grobs = plot_model(mod.rand, type = 're'))
```



```
plot_model(mod.rand, type = 'diag')
```



Random effects: How many levels? Random vs. Fixed?

- Block only has 3 levels (and 5/6 for Year/Paddock), yet we're trying to estimate a variance
- Should we used a fixed effect instead?

Philosophical approach

- **Fixed** = particular groups you want to compare to each other
- Experimental treatments; the 8 Hawaiian islands; the 6 species in your analysis

- **Random** = you need to quantify variation among groups in general
- Many sites, years, individuals that have multiple observations

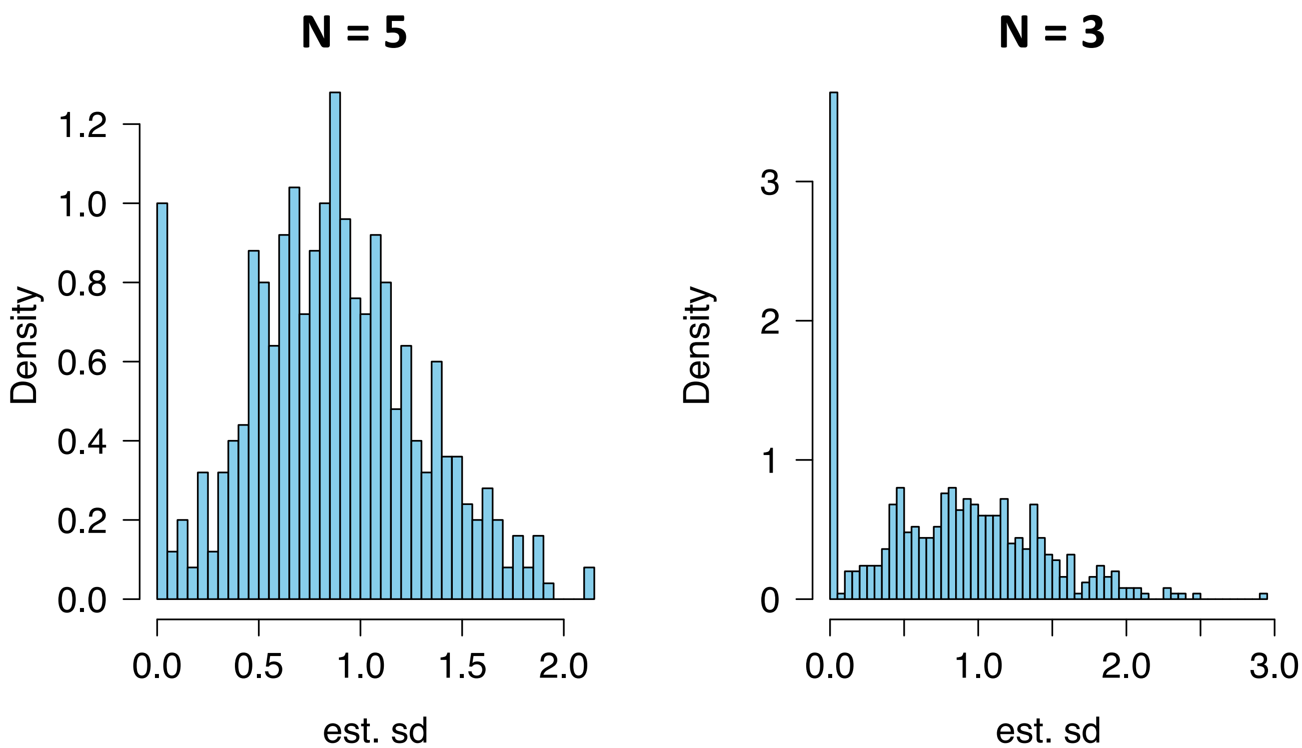
- Controlling for non-independence: can do this with either, but often random makes more sense

- Can't have nested fixed effects

Random effects: How many levels? Random vs. Fixed?

Pragmatic approach

- How many groups were sampled?
- For small samples, more likely to get a variance estimate of zero



5% **18%**

```r
mod.fix = lmer(richness ~ Treatment + (1|Year) + (1|TransectID) + Block + (1|P
addock), data = butterfly)

summary(mod.fix)

## Random effects:
##  Groups     Name        Variance Std.Dev.
##  TransectID (Intercept) 0.000    0.000
##  Paddock    (Intercept) 0.709    0.842
##  Year       (Intercept) 2.494    1.579
##  Residual               2.513    1.585
## Number of obs: 90, groups:  TransectID, 18; Paddock, 6; Year, 5
##
## Fixed effects:
##              Estimate Std. Error     df t value Pr(>|t|)
## (Intercept)     4.489      1.041  4.890    4.31    0.008 **
## Treatmentlow    1.422      0.764  2.000    1.86    0.204
## BlockB         -0.700      0.936  2.000   -0.75    0.533
## BlockC         -1.033      0.936  2.000   -1.10    0.385
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

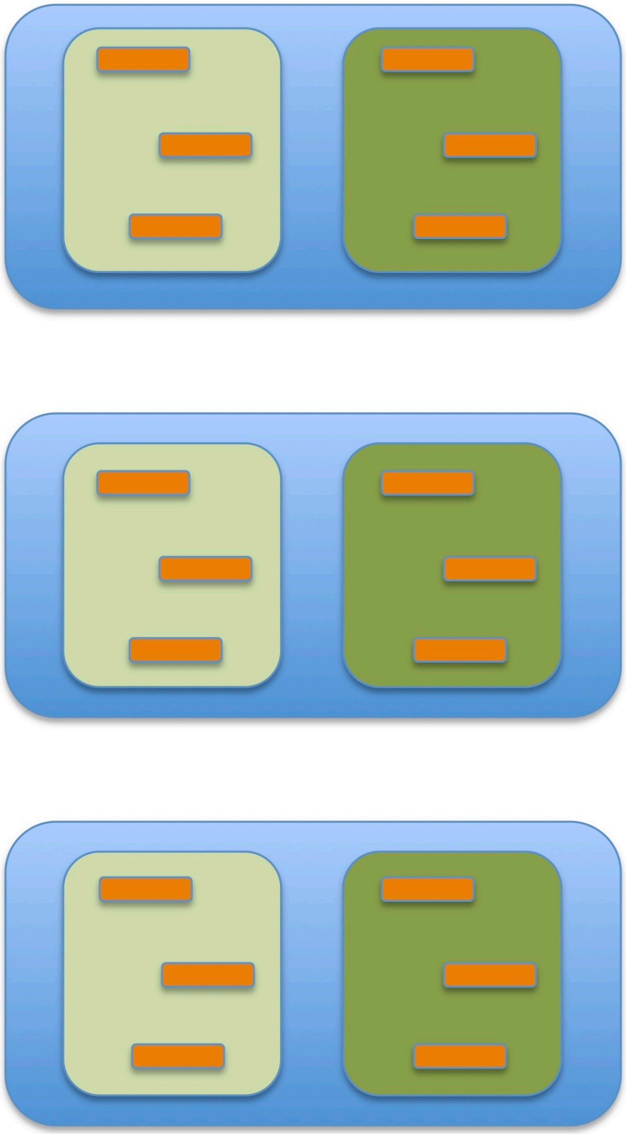
Testing the treatment effect – methods can matter for small effective sample size

```
anova(mod, ddf = "Kenward-Roger")

## Analysis of Variance Table of type 3  with  Kenward-Roger
## approximation for degrees of freedom
##           Sum Sq Mean Sq NumDF DenDF F.value Pr(>F)
## Treatment   13.8    13.8     1     2    4.24   0.18
```

```
anova(mod, ddf = "Satterthwaite")

## Analysis of Variance Table of type 3  with  Satterthwaite
## approximation for degrees of freedom
##           Sum Sq Mean Sq NumDF DenDF F.value Pr(>F)
## Treatment   13.8    13.8     1  5.33    5.49  0.063 .
```

```
mod.null = lmer(richness ~ 1 + (1|Year) + (1|TransectID) + (1|Block) +
(1|Paddock), data = butterfly, REML = FALSE)

PBmodcomp(mod, mod.null)

## Parametric bootstrap test; time: 50.02 sec; samples: 1000 extremes: 140;
## large : richness ~ Treatment + (1 | Year) + (1 | TransectID) + (1 | Block)
+
##     (1 | Paddock)
## small : richness ~ 1 + (1 | Year) + (1 | TransectID) + (1 | Block) +
##     (1 | Paddock)
##        stat df p.value
## LRT    3.79  1   0.052 .
## PBtest 3.79      0.141
```

Testing the treatment effect



Block

Paddock = Treatment

Transect

```r
library(MuMIn)
AICc(mod)

## [1] 373.6

AICc(mod.null)

## [1] 375
```

Random treatment*year interaction: multivariate random effects

- Exploratory data analysis, and logic, suggests the treatment effect may vary over time
- How to test Treatment*Year interaction
- Could change Year to fixed effect, then Treatment*Year = 10 parms
- Or could do interaction between Treatment (fixed) and Year (random)

Random treatment*year interaction: multivariate random effects

```
mod.rand = lmer(richness ~ Treatment + (1|Year) + (1|TransectID) + (1|Block) +
(1|Paddock), data = butterfly)
```

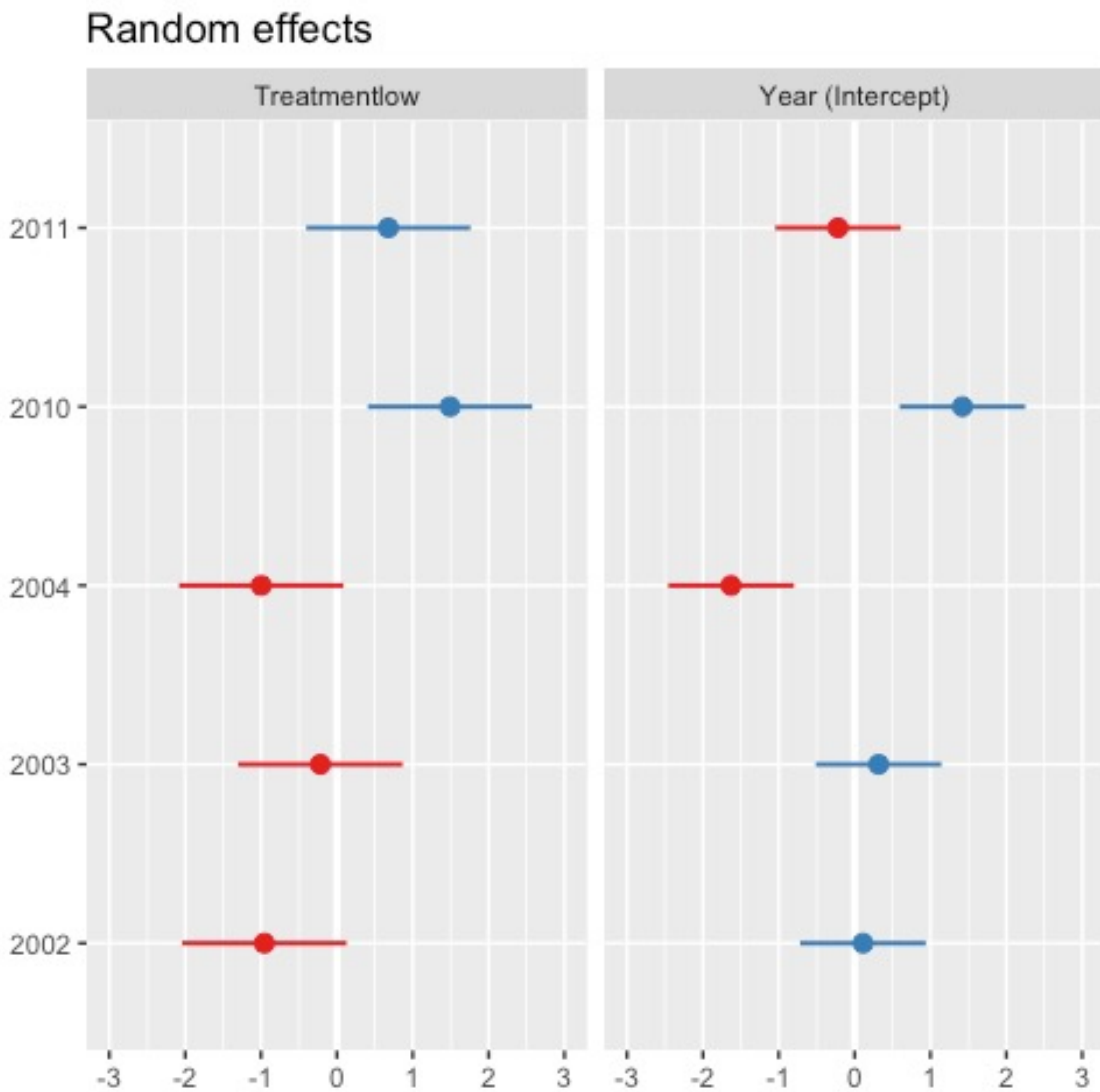
```
mod.inter = lmer(richness ~ Treatment + (Treatment|Year) + (1|TransectID) +
(1|Block) + (1|Paddock), data = butterfly)
```

```
VarCorr(mod.inter)

##  Groups     Name         Std.Dev. Corr
##  TransectID (Intercept)  0.00e+00
##  Paddock    (Intercept)  7.56e-01
##  Year       (Intercept)  1.18e+00
##             Treatmentlow 1.23e+00 0.50
##  Block      (Intercept)  1.78e-07
##  Residual                1.47e+00
```

- Now the Intercept and the Treatment effect both vary randomly by Year

- Corr = whether the Year-specific Intercept tends to be correlated with the Year-specific Treatment effect

```
plot_model(mod.inter, type = 're')
```



Random treatment*year interaction: multivariate random effects

- If there are multiple parameters that vary by group, these are assumed to come from a **multivariate normal distribution**

$$\vec{\mu} = \begin{bmatrix} \mu_I \\ \mu_T \end{bmatrix}, \Sigma = \begin{bmatrix} \sigma_I^2 & \sigma_{IT} \\ \sigma_{IT} & \sigma_T^2 \end{bmatrix}$$

- A multivariate distribution generates a **random vector**
- The vector has a **mean** and a **variance-covariance matrix**
- The matrix determines the variability and correlation among elements of the vector
- $\boldsymbol{\sigma_I^2}$ is the random effects variance for Intercept
- $\boldsymbol{\sigma_T^2}$ is the random effects variance for Treatmentlow
- $\boldsymbol{\sigma_{IT}}$ is the **covariance** among the two effects

Random treatment*year interaction: multivariate random effects

- If there are multiple parameters that vary by group, these are assumed to come from a **multivariate normal distribution**

$$\vec{\mu} = \begin{bmatrix} \mu_I \\ \mu_T \end{bmatrix}, \Sigma = \begin{bmatrix} \sigma_I^2 & \sigma_{IT} \\ \sigma_{IT} & \sigma_T^2 \end{bmatrix}$$

- The covariance can be rewritten in terms of the correlation: $\boldsymbol{\sigma_{IT} = \rho_{IT}\sigma_I\sigma_T}$

- The matrix can be written like this instead

$$\begin{bmatrix} \sigma_I^2 & \rho_{IT} \\ \rho_{IT} & \sigma_T^2 \end{bmatrix}$$

- This is what lmer returns.

- Correlation of 0.5 means years with high overall butterfly richness also tend to have a higher treatment effect

- The ranef() are centered around 0; the vector means are given by fixef()

Random treatment*year interaction: multivariate random effects

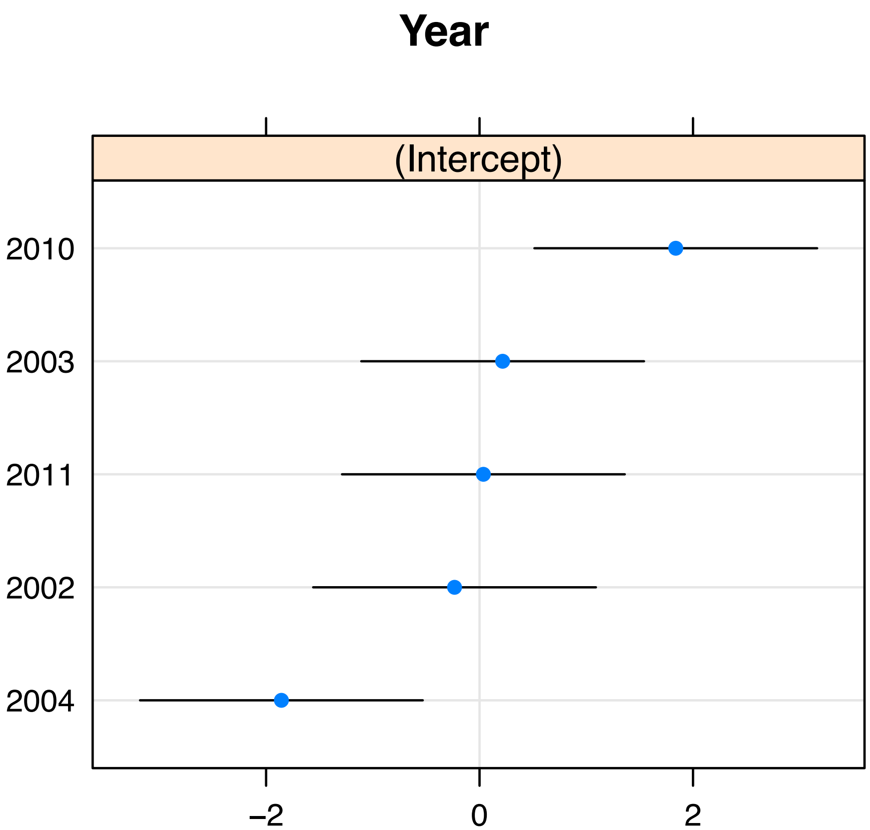
- You may want to assume the random effects are **uncorrelated**

```
mod.nocorr = lmer(richness ~ Treatment + (1|Year) + (1|Year:Treatment)
+ (1|TransectID) + (1|Block) + (1|Paddock), data = butterfly)
```

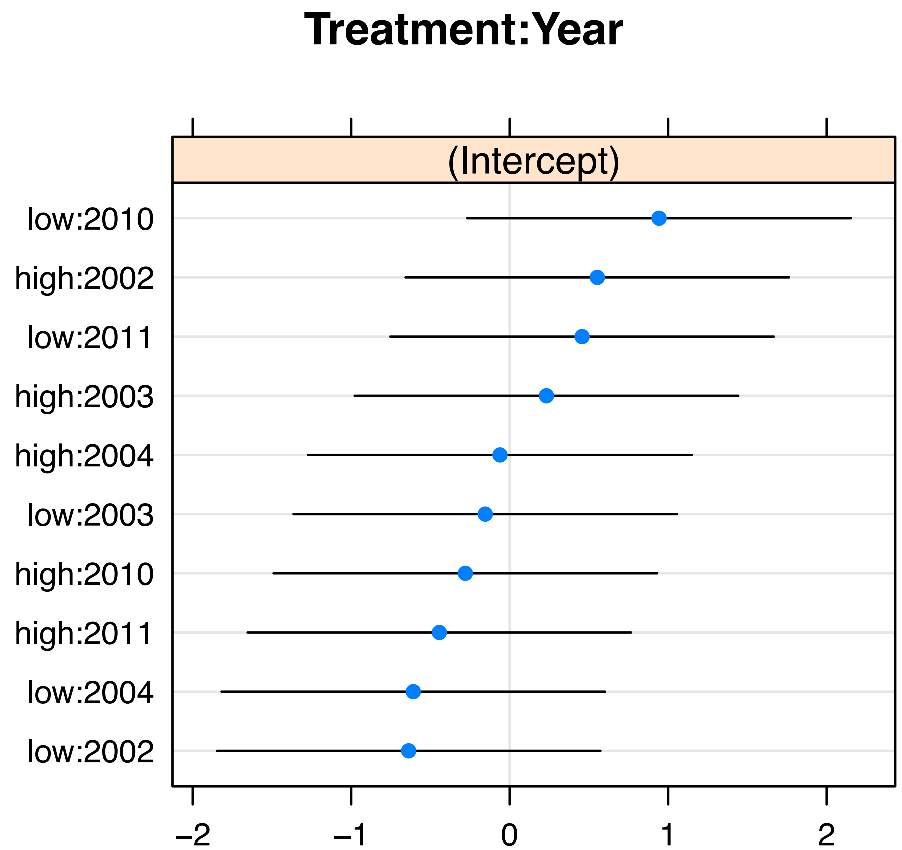
- Variation among years; and then variation among Year:Treatment combinations

- Doesn't pay attention to whether effects come from the same year

```
## Random effects:
##  Groups         Name        Variance Std.Dev.
##  TransectID     (Intercept) 0.000    0.000
##  Year:Treatment (Intercept) 0.753    0.868
##  Paddock        (Intercept) 0.571    0.756
##  Year           (Intercept) 2.136    1.461
##  Block          (Intercept) 0.000    0.000
##  Residual                   2.174    1.475
## Number of obs: 90, groups:
## TransectID, 18; Year:Treatment, 10; Paddock, 6; Year, 5; Block, 3
```



Year
(Intercept)
2010
2003
2011
2002
2004
−2
0
2



Treatment:Year
(Intercept)
low:2010
high:2002
low:2011
high:2003
high:2004
low:2003
high:2010
high:2011
low:2004
low:2002
−2
−1
0
1
2

Testing random effects

- Often there's no need
- Fixed*random interaction is something you probably want to test
- The rules of thumb are not the same: **LRT tends to be conservative**
- Variance parameter cannot go below 0
- The upshot that LRT and AIC are at least not going to give spurious results

Testing random effects

```r
mod = lmer(richness ~ Treatment + (1|Year) + (1|TransectID) + (1|Block) +
(1|Paddock), data = butterfly)
mod.inter = lmer(richness ~ Treatment + (Treatment|Year) + (1|TransectID) +
(1|Block) + (1|Paddock), data = butterfly)
```

```r
anova(mod, mod.inter, test = "Chisq")

## refitting model(s) with ML (instead of REML)

## Data: butterfly
## Models:
## mod: richness ~ Treatment + (1 | Year) + (1 | TransectID) + (1 | Block) +
## mod:     (1 | Paddock)
## mod.inter: richness ~ Treatment + (Treatment | Year) + (1 | TransectID) +
## mod.inter:     (1 | Block) + (1 | Paddock)
##           Df AIC BIC logLik deviance Chisq Chi Df Pr(>Chisq)
## mod        7 372 390   -179      358
## mod.inter  9 368 391   -175      350   7.8      2       0.02 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Testing random effects

```
PBmodcomp(mod.inter, mod)

## Parametric bootstrap test; time: 64.27 sec; samples: 1000 extremes: 4;
## Requested samples: 1000 Used samples: 997 Extremes: 4
## large : richness ~ Treatment + (Treatment | Year) + (1 | TransectID) +
##     (1 | Block) + (1 | Paddock)
## small : richness ~ Treatment + (1 | Year) + (1 | TransectID) + (1 | Block) +
##     (1 | Paddock)
##          stat df p.value
## LRT       7.8  2   0.020 *
## PBtest    7.8      0.005 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
hightreat.predicted = fixef(mod.inter)["(Intercept)"] + ranef(mod.inter, whichel = "Year")$Year[,"(Intercept)"]

lowtreat.predicted = fixef(mod.inter)["(Intercept)"] + ranef(mod.inter, whichel = "Year")$Year[,"(Intercept)"] + fixef(mod.inter)["Treatmentlow"] + ranef(mod.inter, whichel = "Year")$Year[,"Treatmentlow"]

plot(hightreat.predicted ~ levels(butterfly$Year), pch = 19, col = 'blue', ylim = c(min(c(hightreat.predicted, lowtreat.predicted)), max(c(hightreat.predicted, lowtreat.predicted))))
points(lowtreat.predicted ~ levels(butterfly$Year), pch = 19, col = 'red')
```