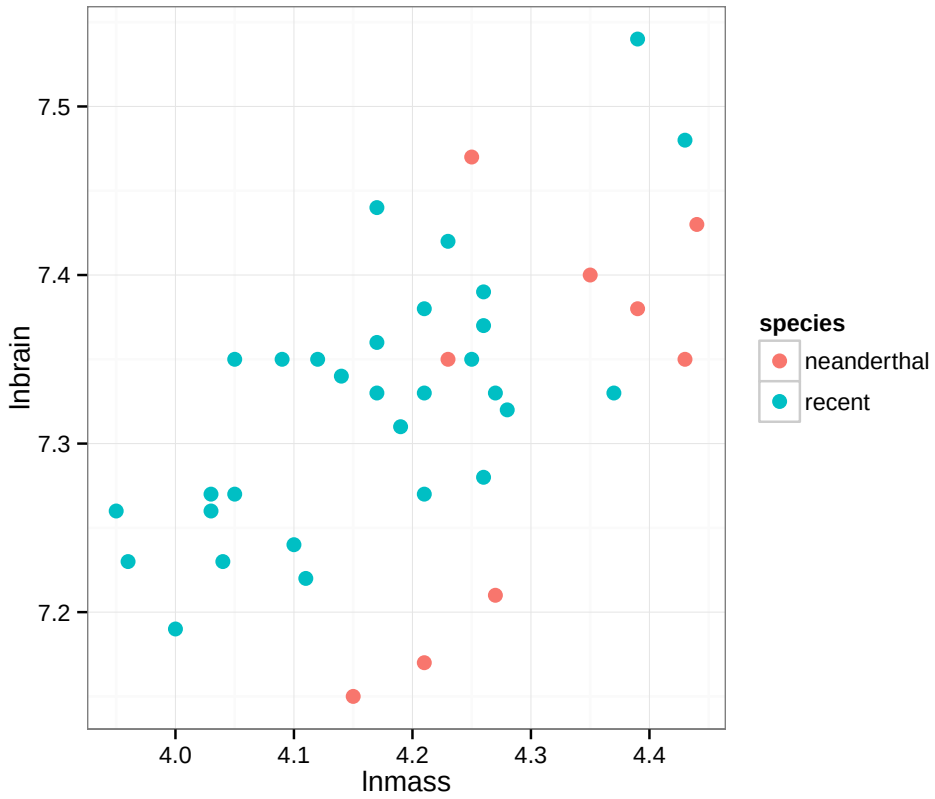
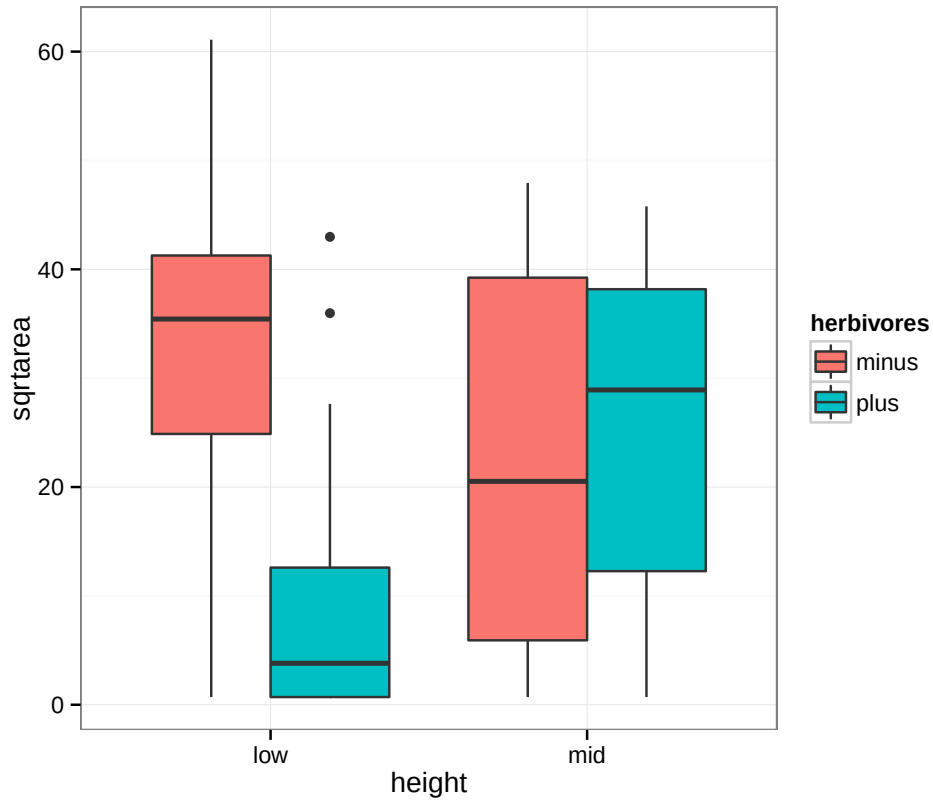
The General Linear Model & Interaction Effects

Problem: How Do you Evalute a Categorical Predictor in the Presence of a Continuous Predictor?



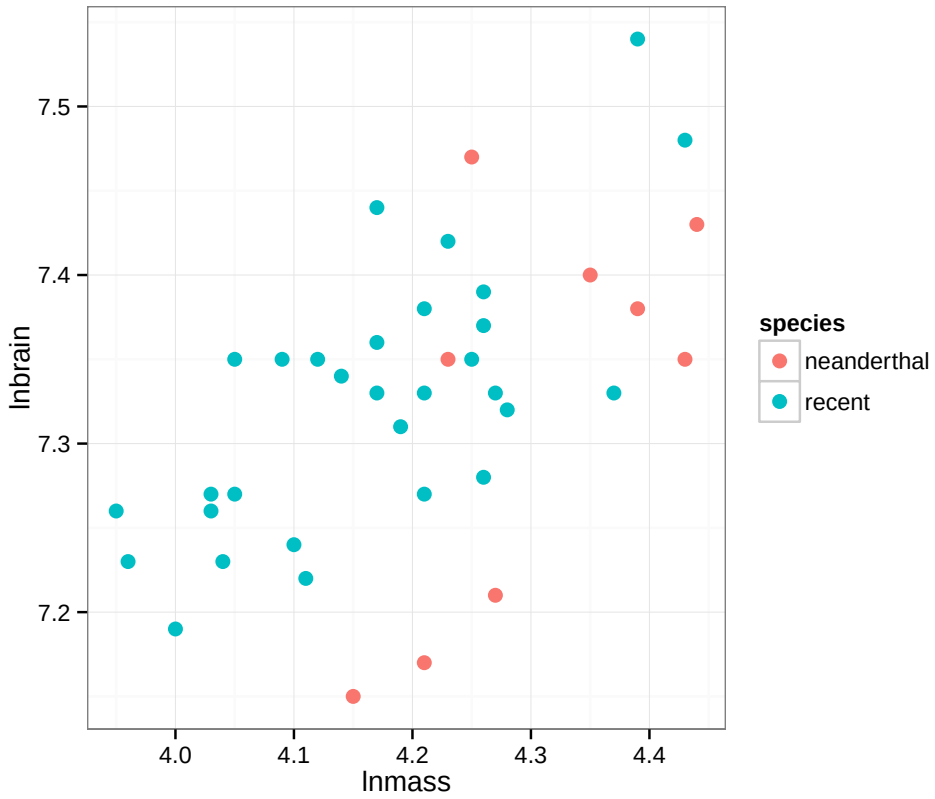
Problem: What if Categorical Predictors are Not Additive?



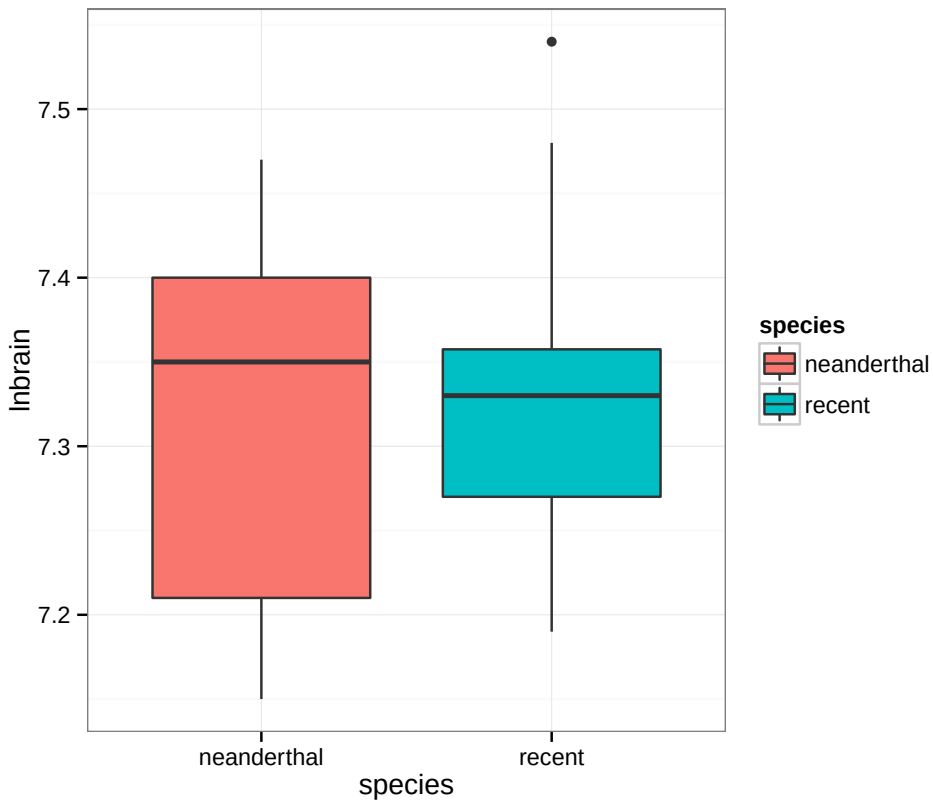
Neanderthals and the General Linear Model

How big was their brain?

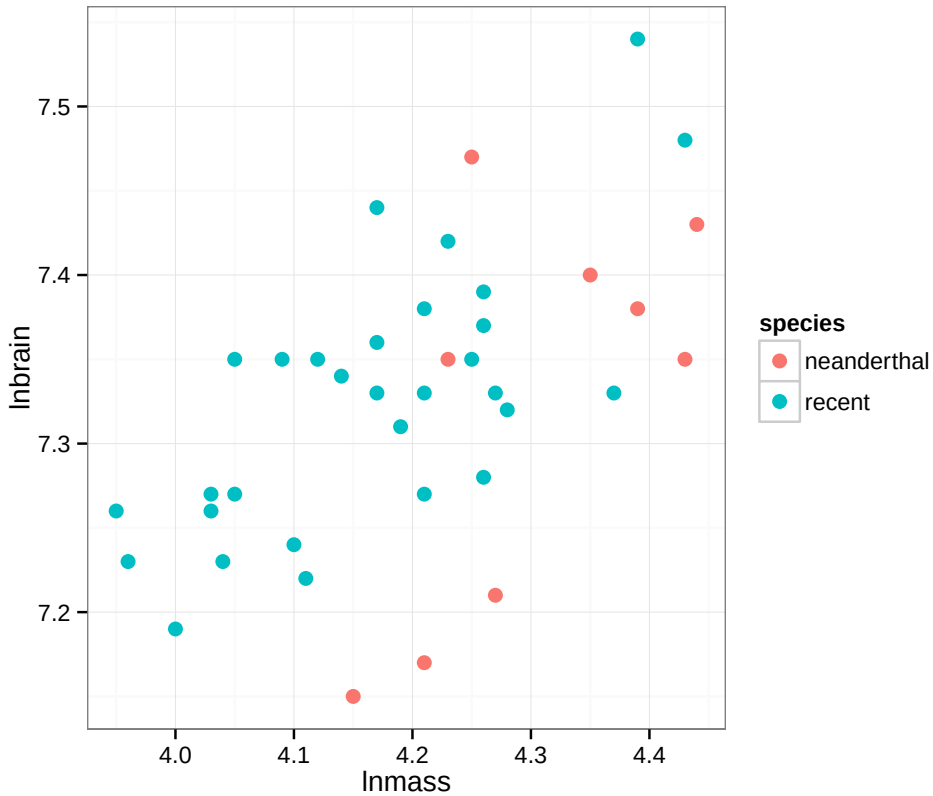
Problem: How Do you Evalute a Categorical Predictor in the Presence of a Continuous Predictor?



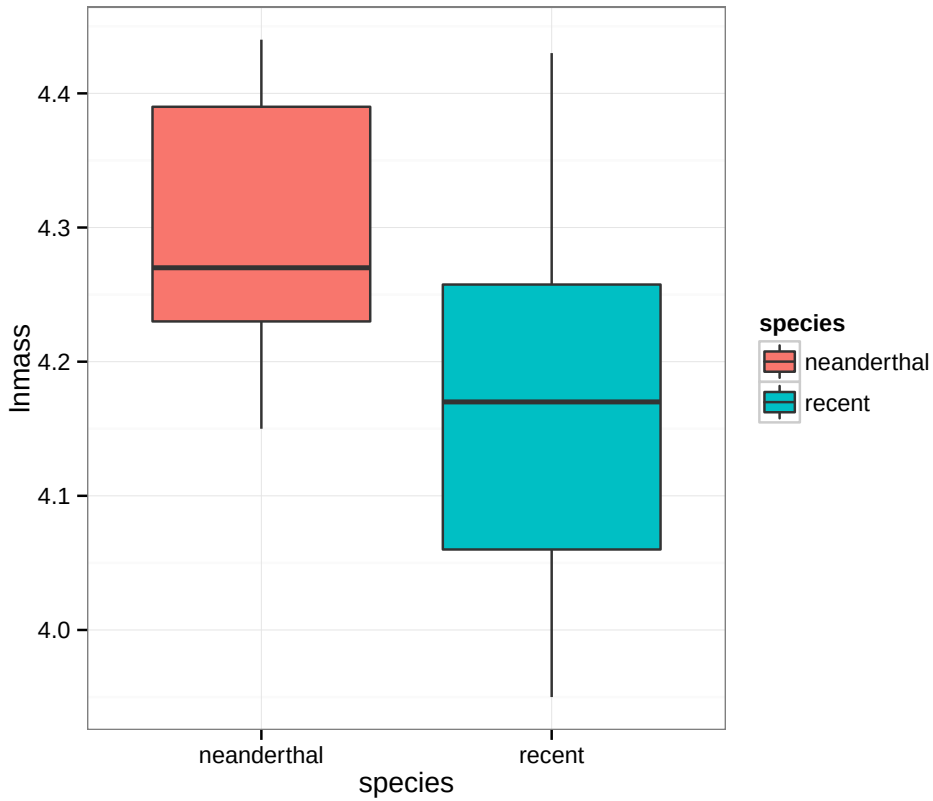
The Means Look the Same...



But there appears to be a Relationship Between Body and Brain Mass



And Mean Body Mass is Different

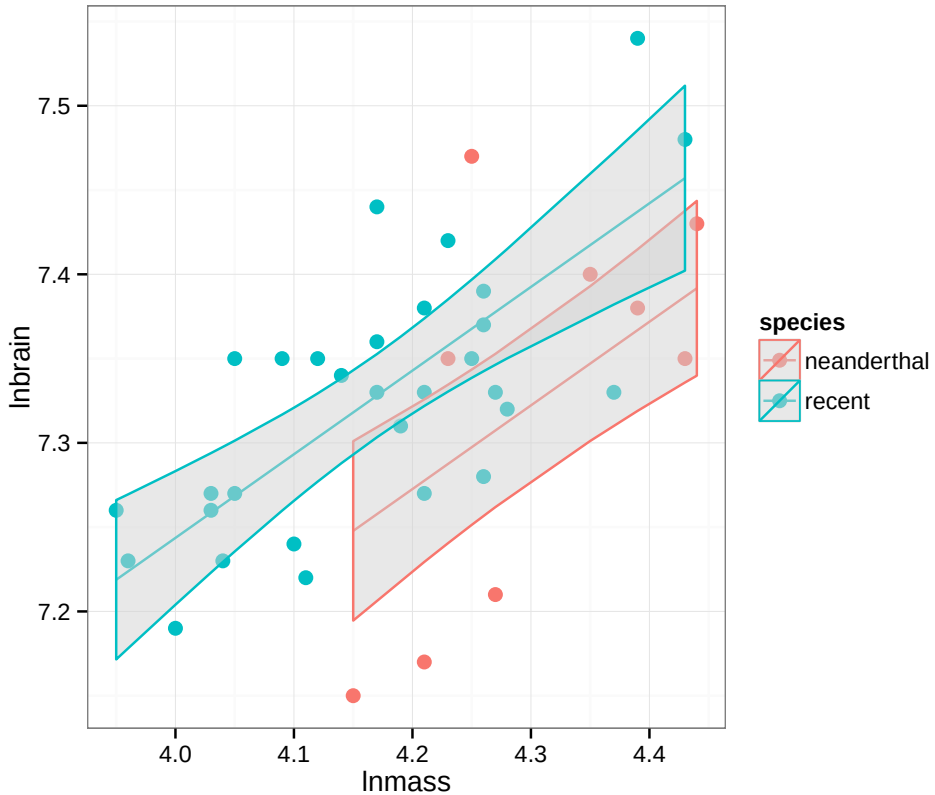


The General Linear Model

$$\boldsymbol{Y} = \boldsymbol{\beta X} + \boldsymbol{\epsilon}$$

- This equation is huge. X can be anything - categorical, continuous, etc.
- We can control for the effect of a covariate - i.e., ANCOVA
- Type of SS matters, as 'covariate' is de facto 'unbalanced'

Analysis of Covariance (control for a covariate)



ANCOVA: Evaluate a categorical effect(s), controlling for a *covariate* (parallel lines)

Groups modify the *intercept*.

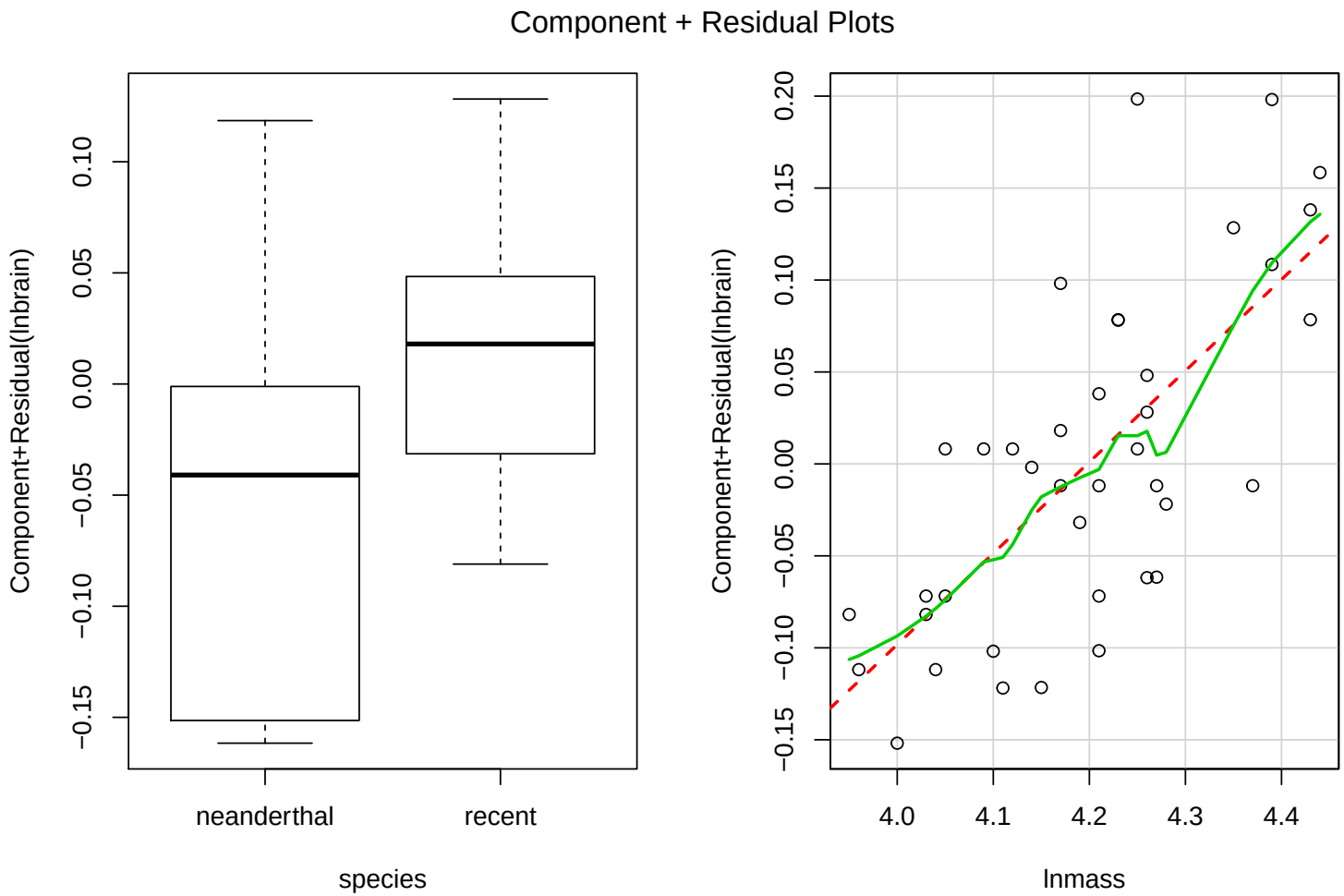
Exercise: Fit like a cave man

- Fit a model that will describe brain size from this data
- Does species matter? Compare type I and type II SS results
- Use Component-Residual plots to evaluate results

Type of SS Matters

```
# Analysis of Variance Table
#
# Response: lnbrain
#           Df Sum Sq Mean Sq F value  Pr(>F)
# species    1 0.0001  0.0001    0.01    0.91
# lnmass     1 0.1300  0.1300   29.28 4.3e-06
# Residuals 36 0.1599  0.0044

# Anova Table (Type II tests)
#
# Response: lnbrain
#           Sum Sq Df F value  Pr(>F)
# species   0.0276  1     6.2   0.017
# lnmass    0.1300  1    29.3 4.3e-06
# Residuals 0.1599 36
```

Species Effect



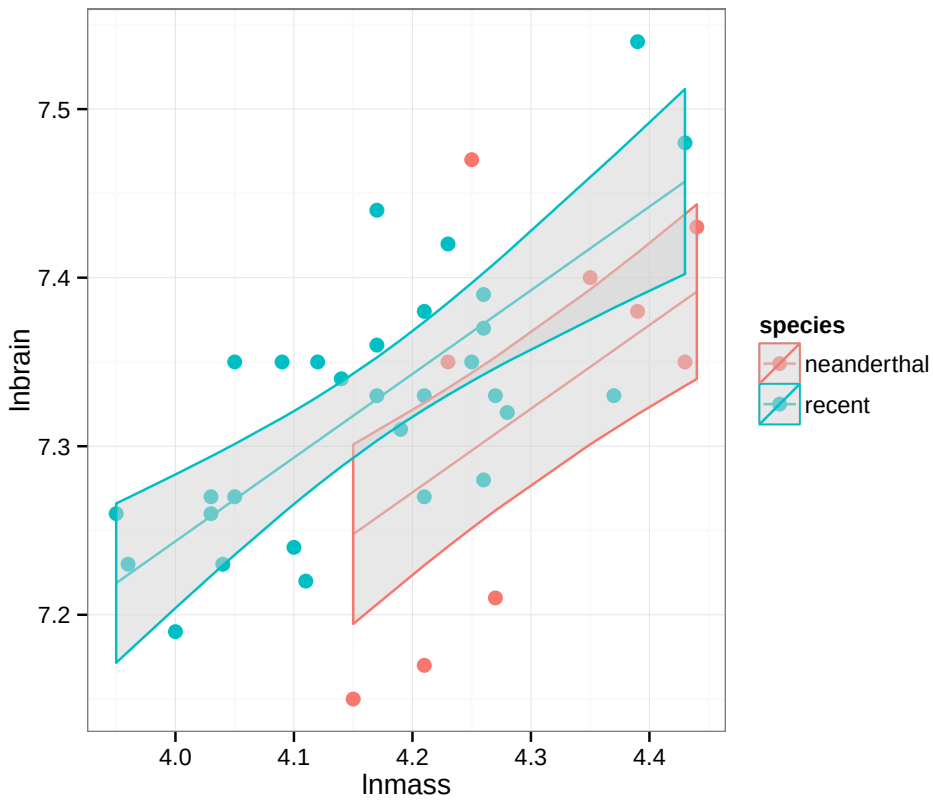
Species Effect

```
summary(neand_lm)$coefficients

#                Estimate Std. Error t value  Pr(>|t|)
# (Intercept)     5.18807    0.39526  13.126 2.736e-15
# speciesrecent   0.07028    0.02822   2.491 1.749e-02
# lnmass          0.49632    0.09173   5.411 4.262e-06

summary(neand_lm)$r.squared

# [1] 0.4486
```

Species Effect

```
contrast(neand_lm,
         list(species="neanderthal", lnmass=mean(neand$lnmass)),
         list(species="recent", lnmass=mean(neand$lnmass)),
         type="average")

# lm model parameter contrast
#
#   Contrast    S.E.   Lower    Upper     t df Pr(>|t|)
# 1 -0.07028 0.02822 -0.1275 -0.01306 -2.49 36   0.0175
```

How to Plot a Fit Model



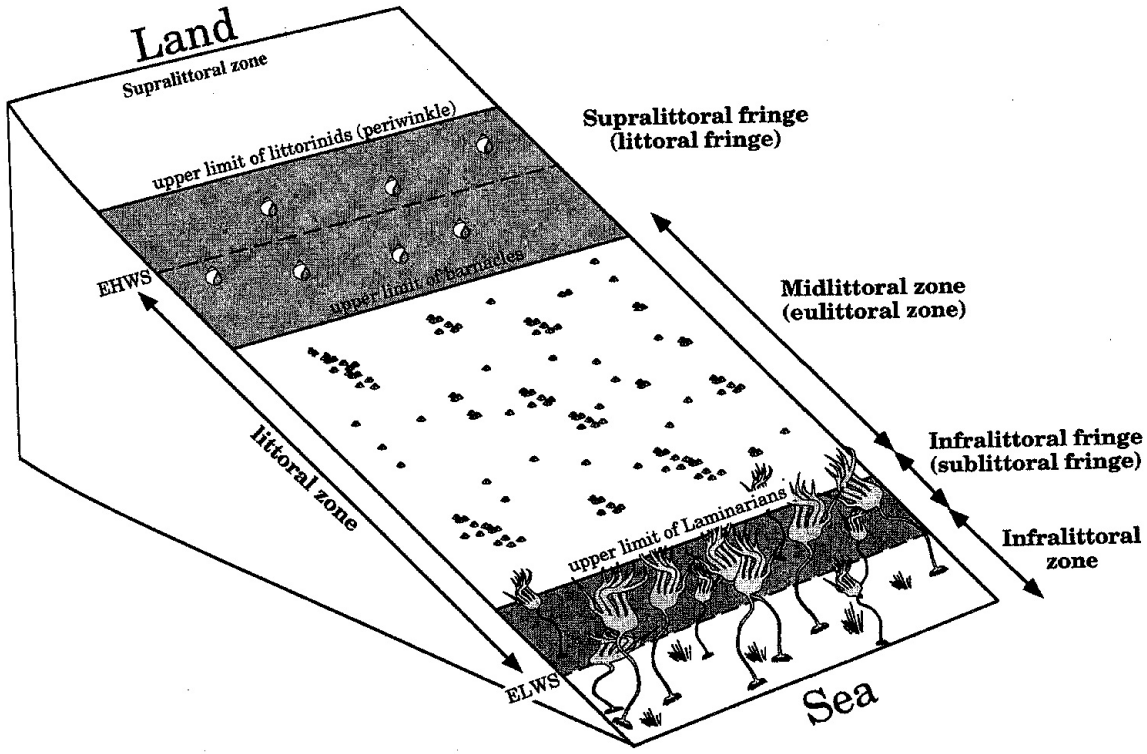
How to Properly Plot a Fit Model

```
neand <- cbind(neand, predict(neand_lm, interval="confidence"))

neand_plot +
  geom_line(data=neand, aes(y=fit)) +
  geom_ribbon(data=neand, aes(ymin=lwr,
                              ymax=upr),
                              fill="lightgrey",
                              alpha=0.5)
```

Intertidal Grazing!

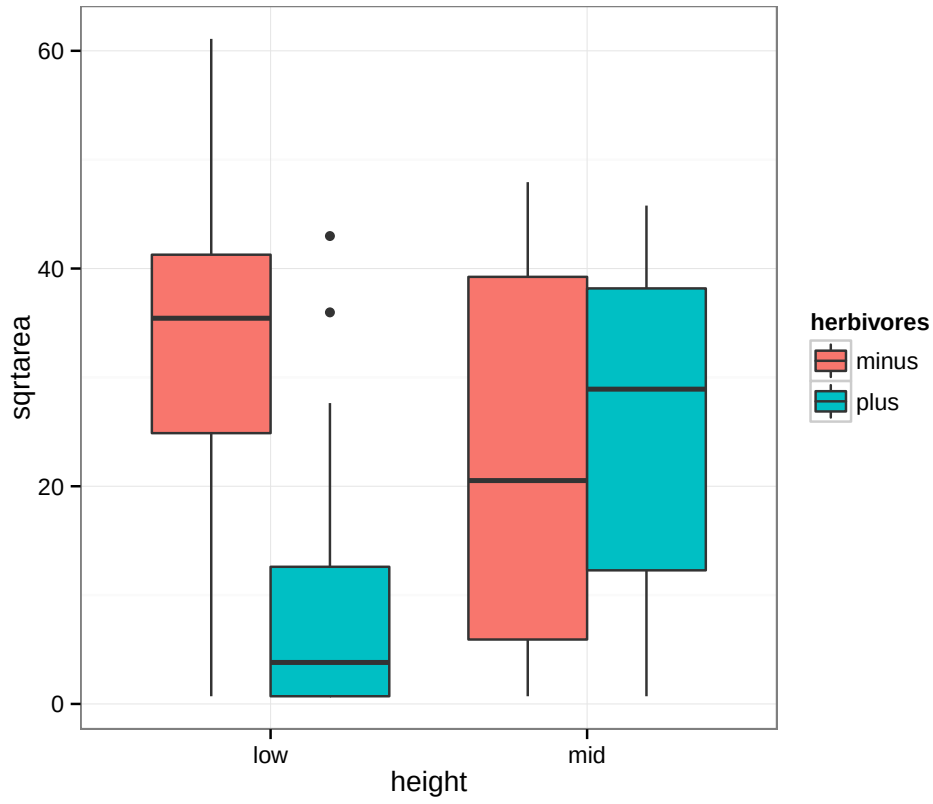
Do grazers reduce algal cover in the intertidal?

Experiment Replicated on Two Ends of a gradient

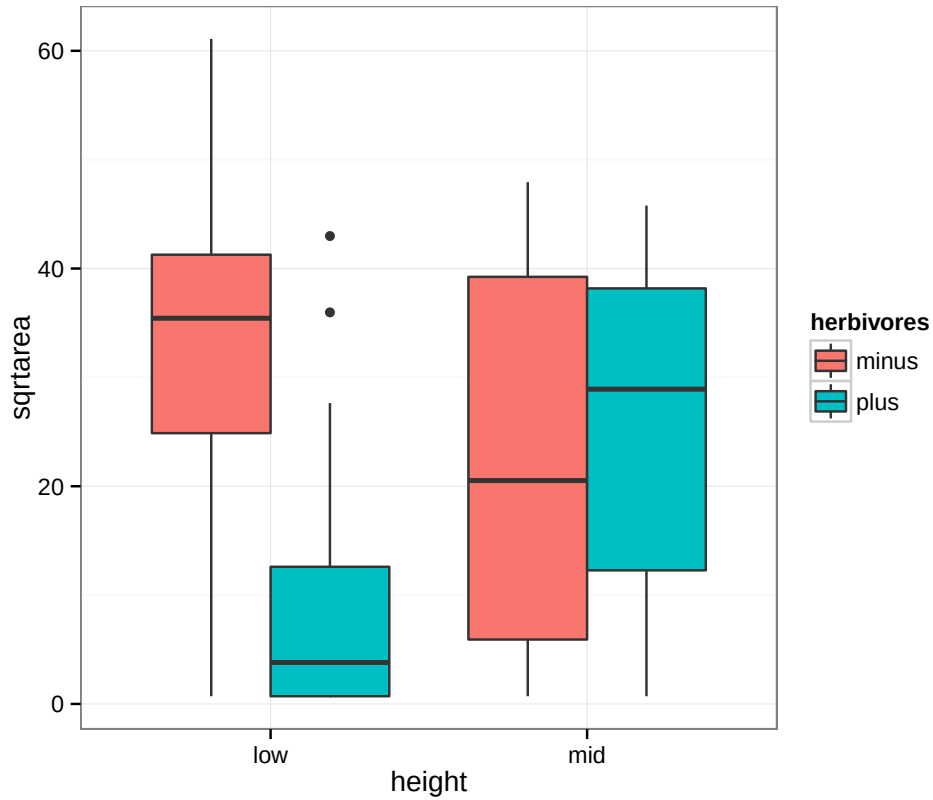


Is this how you would have done it?

Problem: What if Categorical Predictors are Not Additive?



Problem: What if Categorical Predictors are Not Additive?

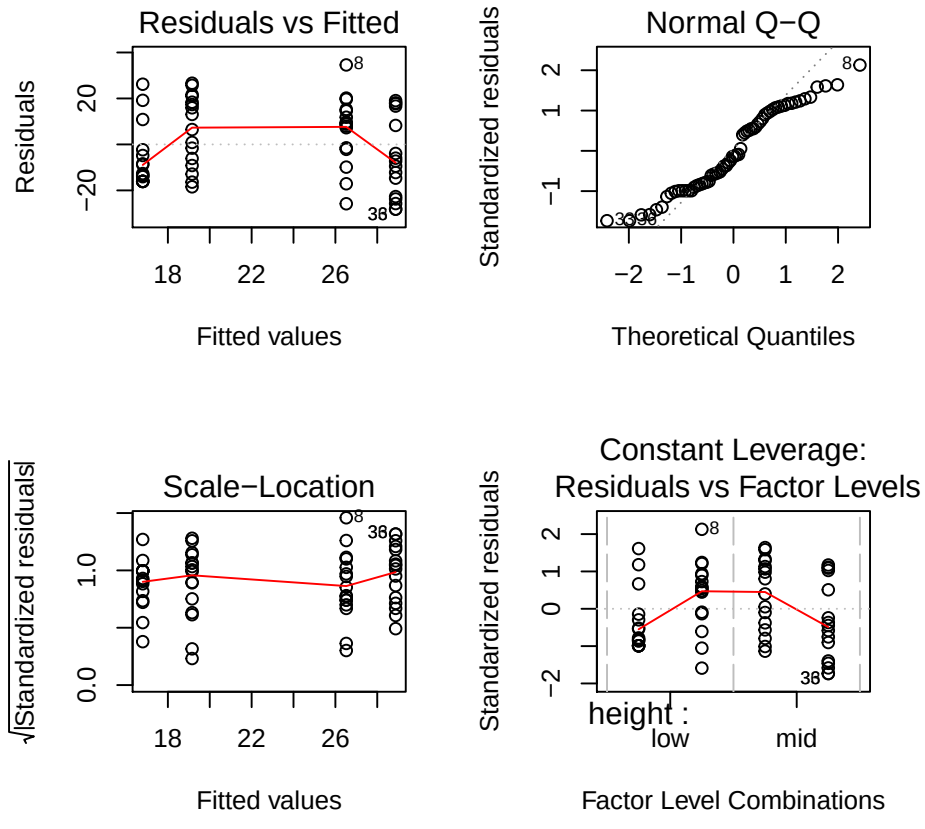


You have replication at the level of block (tide height) and treatment (grazing)

What if you fit a linear model?

```
graze_linear <- lm(sqrtarea ~ height + herbivores, data=algae)
Anova(graze_linear)

# Anova Table (Type II tests)
#
# Response: sqrtarea
#             Sum Sq Df F value Pr(>F)
# height          89  1    0.32  0.573
# herbivores    1512  1    5.46  0.023
# Residuals    16887 61
```

Pattern in Fitted v. Residuals



Nonlinearity!

```
residualPlots(graze_linear, plot=F)

#             Test stat Pr(>|t|)
# height             NA       NA
# herbivores         NA       NA
# Tukey test     -3.317    0.001
```

The General Linear Model

$$\boldsymbol{Y} = \boldsymbol{\beta X} + \boldsymbol{\epsilon}$$

- $\boldsymbol{X}$ can have Nonlinear predictors
- e.g., It can encompass A, B, and A*B

How do you Fit a Multiplicative Effect?

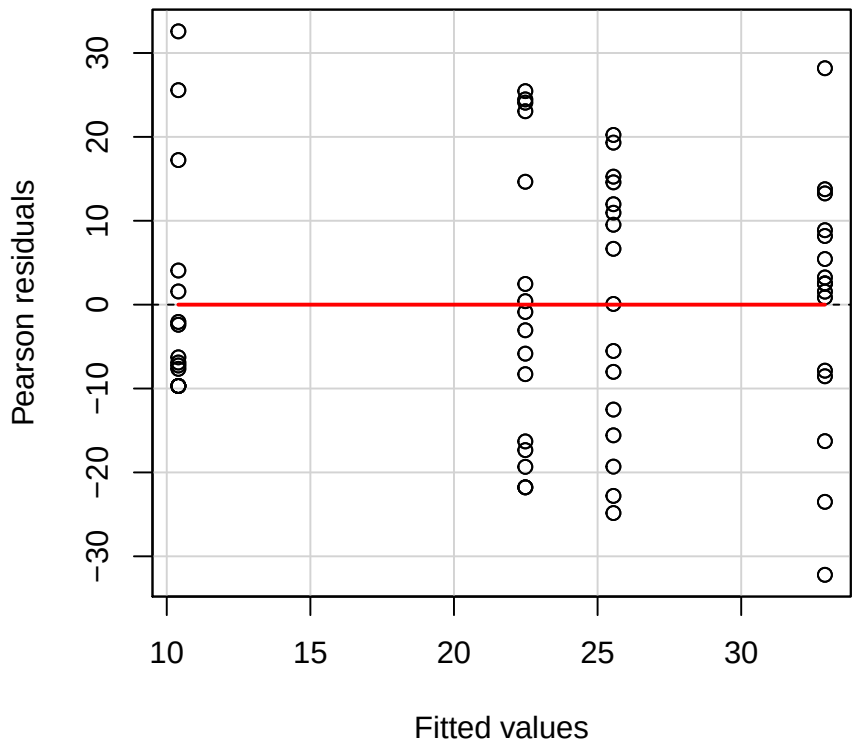
```
graze_int <- lm(sqrtarea ~ height + herbivores + herbivores:height,
                data=algae)
```

How do you Fit a Multiplicative Effect?

```
graze_int <- lm(sqrtarea ~ height + herbivores + herbivores:height,
                data=algae)
```

```
#Or, more compact syntax
graze_int <- lm(sqrtarea ~ height*herbivores, data=algae)
```

No More Pattern in Fitted v. Residuals



F-Tests for Interactions

$$SS_{Total} = SS_A + SS_B + SS_{AB} + SS_{Error}$$

$$SS_{AB} = n \sum_i \sum_j (\bar{Y_{ij}} - \bar{Y_i} - \bar{Y_j} - \bar{Y})^2, \text{ df=(i-1)(j-1)}$$

F-Tests for Interactions

$$SS_{Total} = SS_A + SS_B + SS_{AB} + SS_{Error}$$

$$SS_{AB} = n \sum_i \sum_j (\bar{Y_{ij}} - \bar{Y_i} - \bar{Y_j} - \bar{Y})^2, \text{ df=(i-1)(j-1)}$$

MS = SS/DF, e.g, $MS_W = \frac{SS_W}{n-k}$

$$F = \frac{MS_{AB}}{MS_{Error}} \text{ with DF=(j-1)(k-1),n - 1 - (i-1) - (j-1) - (i-1)(j-1)}$$

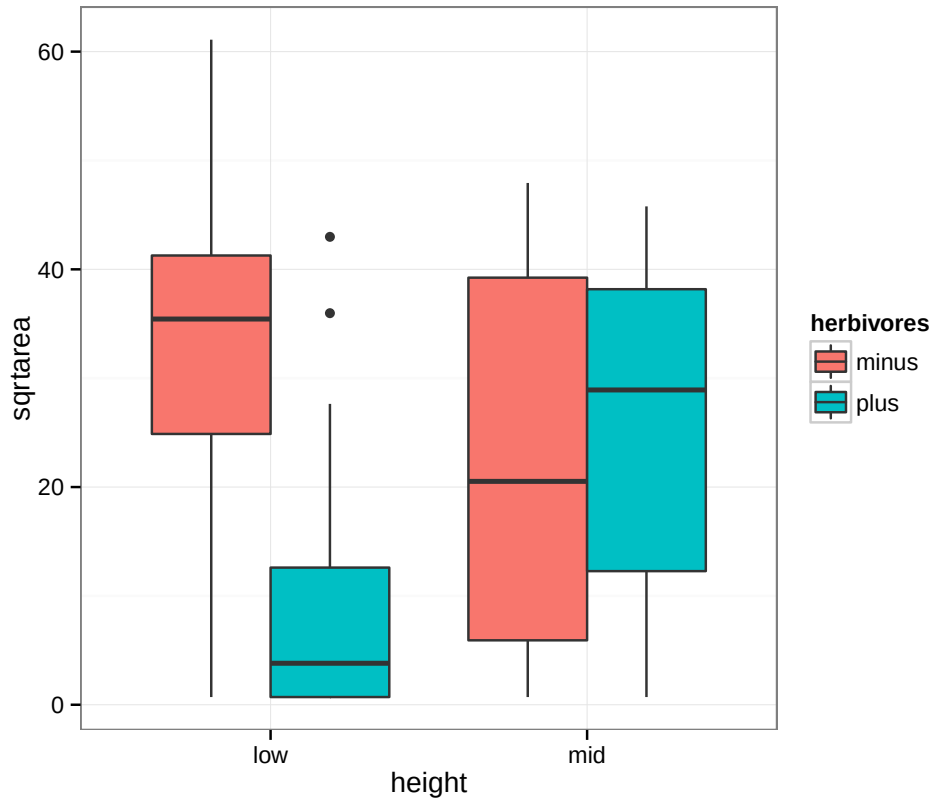
ANOVA shows an Interaction Effect

```
# Anova Table (Type II tests)
#
# Response: sqrtarea
#                    Sum Sq Df F value Pr(>F)
# height                89  1    0.37 0.5431
# herbivores          1512  1    6.36 0.0144
# height:herbivores   2617  1   11.00 0.0015
# Residuals          14271 60
```

What does the Interaction Coefficient Mean?

```
#                            Estimate Std. Error t value
# (Intercept)                   32.91      3.856   8.537
# heightmid                    -10.43      5.453  -1.913
# herbivoresplus               -22.51      5.453  -4.128
# heightmid:herbivoresplus      25.58      7.711   3.317
#                             Pr(>|t|)
# (Intercept)                5.980e-12
# heightmid                  6.052e-02
# herbivoresplus             1.146e-04
# heightmid:herbivoresplus   1.549e-03
```

What does the Interaction Coefficient Mean?



Posthoc with Simple Effects Model

```
algae$int <- with(algae, interaction(height, herbivores))
graze_int2 <- lm(sqrtarea ~ int, data=algae)
#
library(multcomp)
summary(glht(graze_int2, linfct=mcp(int = "Tukey")))
```

Posthoc with Simple Effects Model

```
#
#   Simultaneous Tests for General Linear Hypotheses
#
# Multiple Comparisons of Means: Tukey Contrasts
#
#
# Fit: lm(formula = sqrtarea ~ int, data = algae)
#
# Linear Hypotheses:
#                            Estimate Std. Error t value
# mid.minus - low.minus == 0   -10.43       5.45   -1.91
# low.plus - low.minus == 0    -22.51       5.45   -4.13
# mid.plus - low.minus == 0     -7.36       5.45   -1.35
# low.plus - mid.minus == 0    -12.08       5.45   -2.22
# mid.plus - mid.minus == 0      3.07       5.45    0.56
# mid.plus - low.plus == 0      15.15       5.45    2.78
#                            Pr(>|t|)
# mid.minus - low.minus == 0    0.233
# low.plus - low.minus == 0    <0.001
# mid.plus - low.minus == 0     0.535
# low.plus - mid.minus == 0     0.131
# mid.plus - mid.minus == 0     0.943
# mid.plus - low.plus == 0      0.036
# (Adjusted p values reported -- single-step method)
```