Multiple Predictor Variables: ANOVA

What if you manipulate two factors?

Block 1	Block 2	Block 3	Block 4
A	B	C	D
B	C	D	A
C	D	A	B
D	A	B	C

Randomized Controlled Blocked Design: Design where each treatment only has 1 replicate of a second treatment

What if you manipulate two factors?

Block 1	Block 2	Block 3	Block 4
A	B	C	D
B	C	D	A
C	D	A	B
D	A	B	C

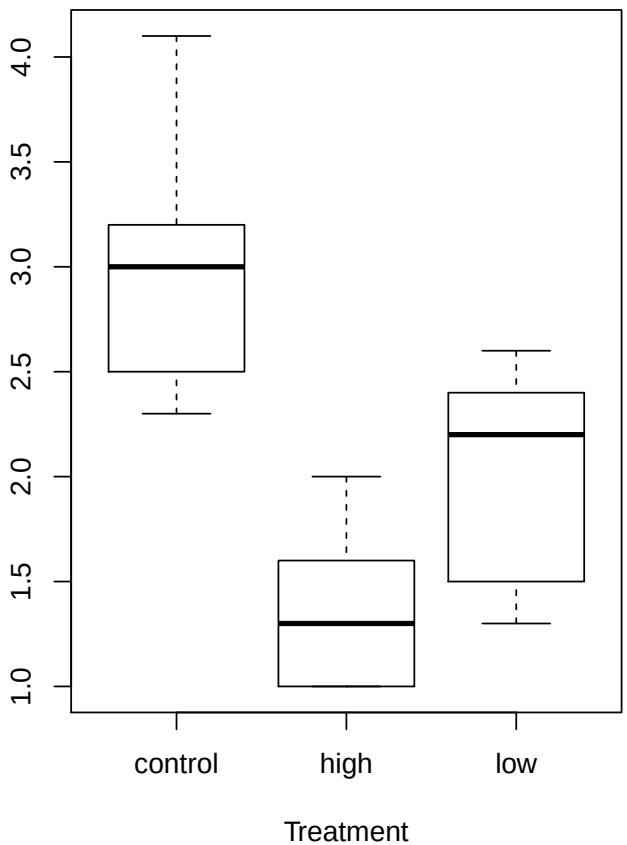
Randomized Controlled Blocked Design: Design where each treatment only has 1 replicate of a second treatment

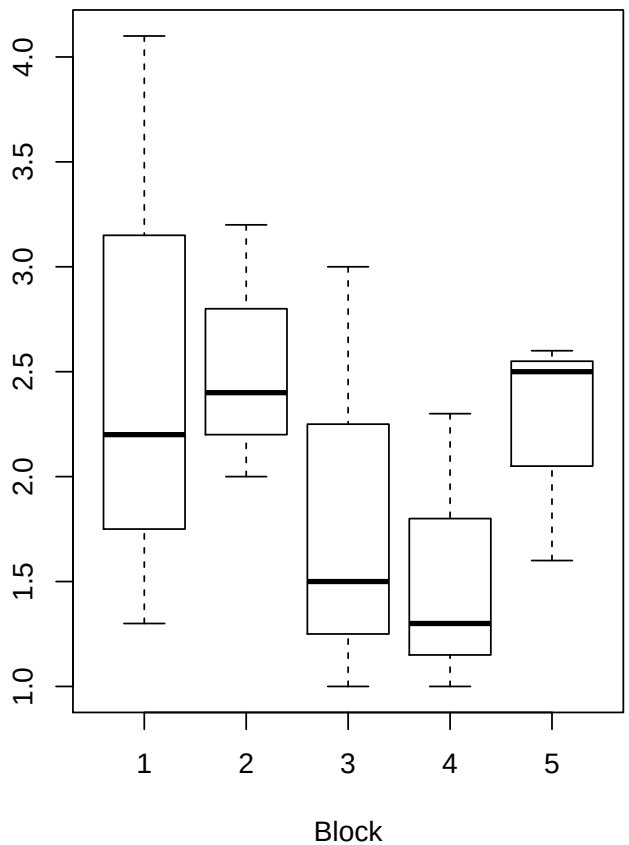
Note: Above is a Latin Squares Design - Every row and column contains one replicate of a treatment.

Effects of Stickleback Density on Zooplankton

Units placed across a lake so that 1 set of each treatment was 'blocked' together

Treatment and Block Effects





Model for Multiway ANOVA/ANODEV

$$y_k = \beta_0 + \sum \beta_i x_i + \sum \beta_j x_j + \epsilon_k$$

$$\epsilon_{ijk} \sim N(0, \sigma^2), \qquad x_i = 0, 1$$

Model for Multiway ANOVA/ANODEV

$$y_k = \beta_0 + \sum \beta_i x_i + \sum \beta_j x_j + \epsilon_k$$

$$\epsilon_{ijk} \sim N(0, \sigma^2), \qquad x_i = 0, 1$$

Or, with matrices...

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

Model for Multiway ANOVA/ANODEV

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

$$\begin{pmatrix} y1 \\ y2 \\ y3 \\ y4 \end{pmatrix} = \begin{pmatrix} \beta_{i1} \\ \beta_{i2} \\ \beta_{j1} \\ \beta_{j2} \end{pmatrix} \begin{pmatrix} 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \end{pmatrix}$$

Model for Multiway ANOVA/ANODEV

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

$$\begin{pmatrix} y1 \\ y2 \\ y3 \\ y4 \end{pmatrix} = \begin{pmatrix} \beta_{i1} \\ \beta_{i2} \\ \beta_{j1} \\ \beta_{j2} \end{pmatrix} \begin{pmatrix} 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \end{pmatrix}$$

hfill

We can have as many groups as we need, so long as there is sufficient replication of each treatment combination.

The 'Treatment Contrast' Model for Multiway ANOVA/ANODEV

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

$$\begin{pmatrix} y1 \\ y2 \\ y3 \\ y4 \end{pmatrix} = \begin{pmatrix} \beta_0 \\ \beta_{i2} \\ \beta_{j2} \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \end{pmatrix}$$

Hypotheses for Multiway ANOVA/ANODEV

TreatmentHo: $\mu_{i1} = \mu i2 = \mu i3 = ...$

Block Ho: $\mu_{j1} = \mu j2 = \mu j3 = ...$

Remember, this can also be stated in terms of β

Sums of Squares for Multiway ANOVA

Factors are Orthogonal and Balanced, so...

- $\text{SST} = \text{SSA} + \text{SSB} + \text{SSR}$
- F-Test using Mean Squares as Before

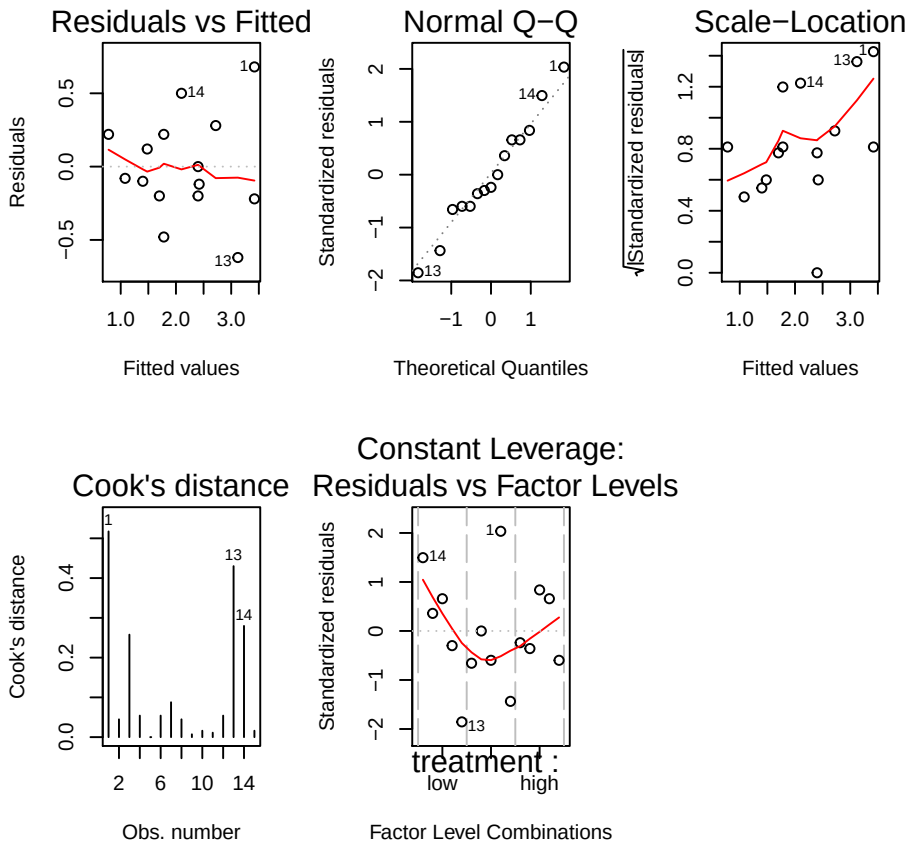
Before we model it, make sure Block is a factor

```
zoop$block <- factor(zoop$block)
```

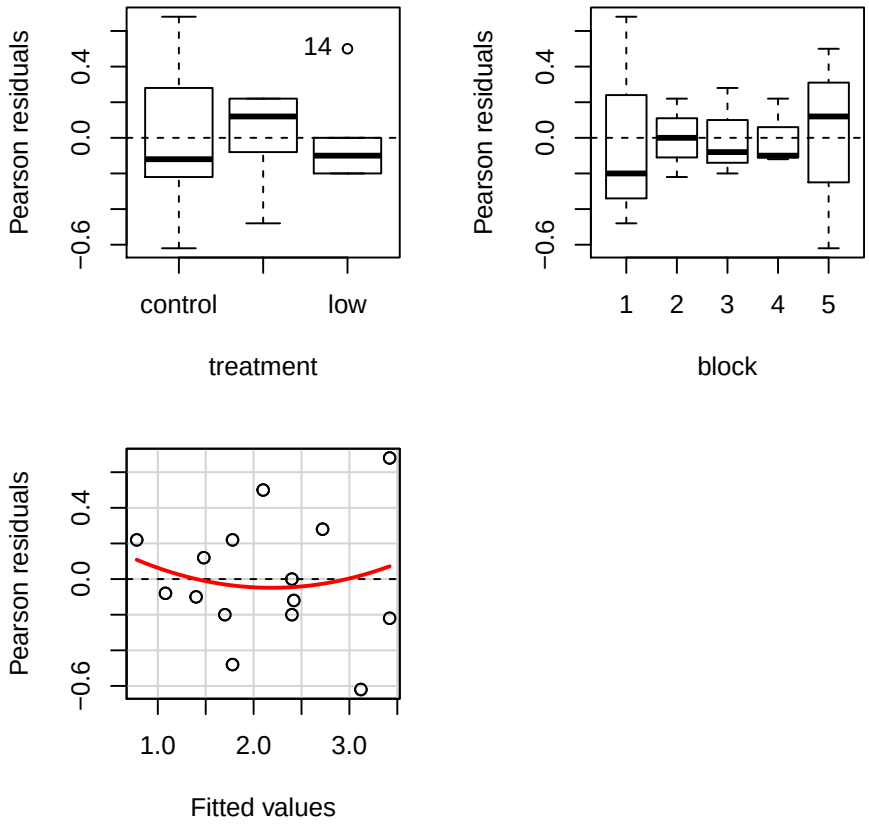
Two-Way ANOVA as a Linear Model

```
zoop_lm <- lm(zooplankton ~ treatment + block, data=zoop)
```

Check Diagnostics



Residuals by Groups and No Non-Additivity



Residuals by Groups and No Non-Additivity

Tukey's Test for Non-Additivity

```
library(car)
residualPlots(zoop_lm)

#             Test stat Pr(>|t|)
# treatment          NA       NA
# block              NA       NA
# Tukey test      0.474    0.635
```

The ANOVA

But first, what are the DF for...
Treatment (with 3 levels)

Block (with 5 blocks)

Residuals (with n=15)

The ANOVA

```
anova(zoop_lm)

# Analysis of Variance Table
#
# Response: zooplankton
#           Df Sum Sq Mean Sq F value Pr(>F)
# treatment  2   6.86    3.43   16.37 0.0015
# block      4   2.34    0.58    2.79 0.1010
# Residuals  8   1.68    0.21
```

Sums of Squares as Model Comparison

Testing SS for a Factor is the same as comparing the residual SS of a model with v. without that factor.

Here is y = intercept versus y = intercept + treatment:

```
zoop_intOnly <- lm(zooplankton ~ 1, data=zoop)
zoop_treatment <- lm(zooplankton ~ treatment , data=zoop)

anova(zoop_intOnly, zoop_treatment)

# Analysis of Variance Table
#
# Model 1: zooplankton ~ 1
# Model 2: zooplankton ~ treatment
#   Res.Df   RSS Df Sum of Sq    F Pr(>F)
# 1     14 10.87
# 2     12  4.02  2      6.86 10.2 0.0025
```

Sums of Squares as Model Comparison

Testing SS for a Factor is the same as comparing the residual SS of a model with v. without that factor.

Here is y = intercept+treatment versus y = intercept + treatment+block:

```
anova(zoop_treatment, zoop_lm)

# Analysis of Variance Table
#
# Model 1: zooplankton ~ treatment
# Model 2: zooplankton ~ treatment + block
#   Res.Df  RSS Df Sum of Sq    F Pr(>F)
# 1     12 4.02
# 2      8 1.68  4      2.34 2.79    0.1
```

Sums of Squares as Model Comparison

Squential model building and SS Calculation is called *T*ype I Sums of Squares

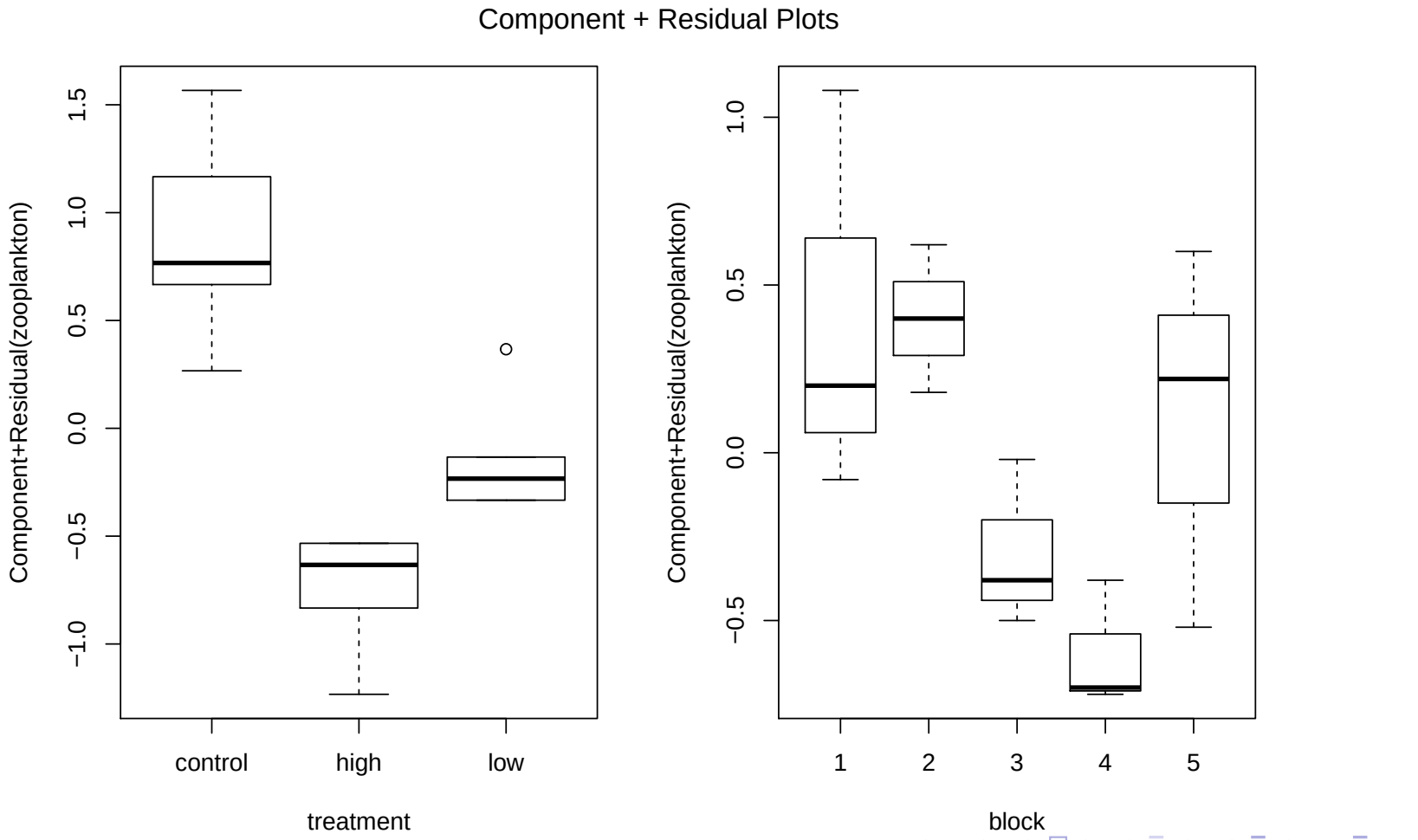
Coefficients via Treatment Contrasts

```
summary(zoop_lm)$coef

#                  Estimate Std. Error    t value  Pr(>|t|)
# (Intercept)    3.420e+00     0.3127  1.094e+01 4.330e-06
# treatmenthigh -1.640e+00     0.2895 -5.665e+00 4.730e-04
# treatmentlow  -1.020e+00     0.2895 -3.524e+00 7.805e-03
# block2         1.039e-15     0.3737  2.781e-15 1.000e+00
# block3        -7.000e-01     0.3737 -1.873e+00 9.795e-02
# block4        -1.000e+00     0.3737 -2.676e+00 2.811e-02
# block5        -3.000e-01     0.3737 -8.027e-01 4.453e-01
```

Unique Effect of Each Treatment

```
crPlots(zoop_lm)
```



Exercise: Bees!

- Load the Bee Gene Expresion Data
- Does bee type or colony matter?
- How much variation does this experiment explain?

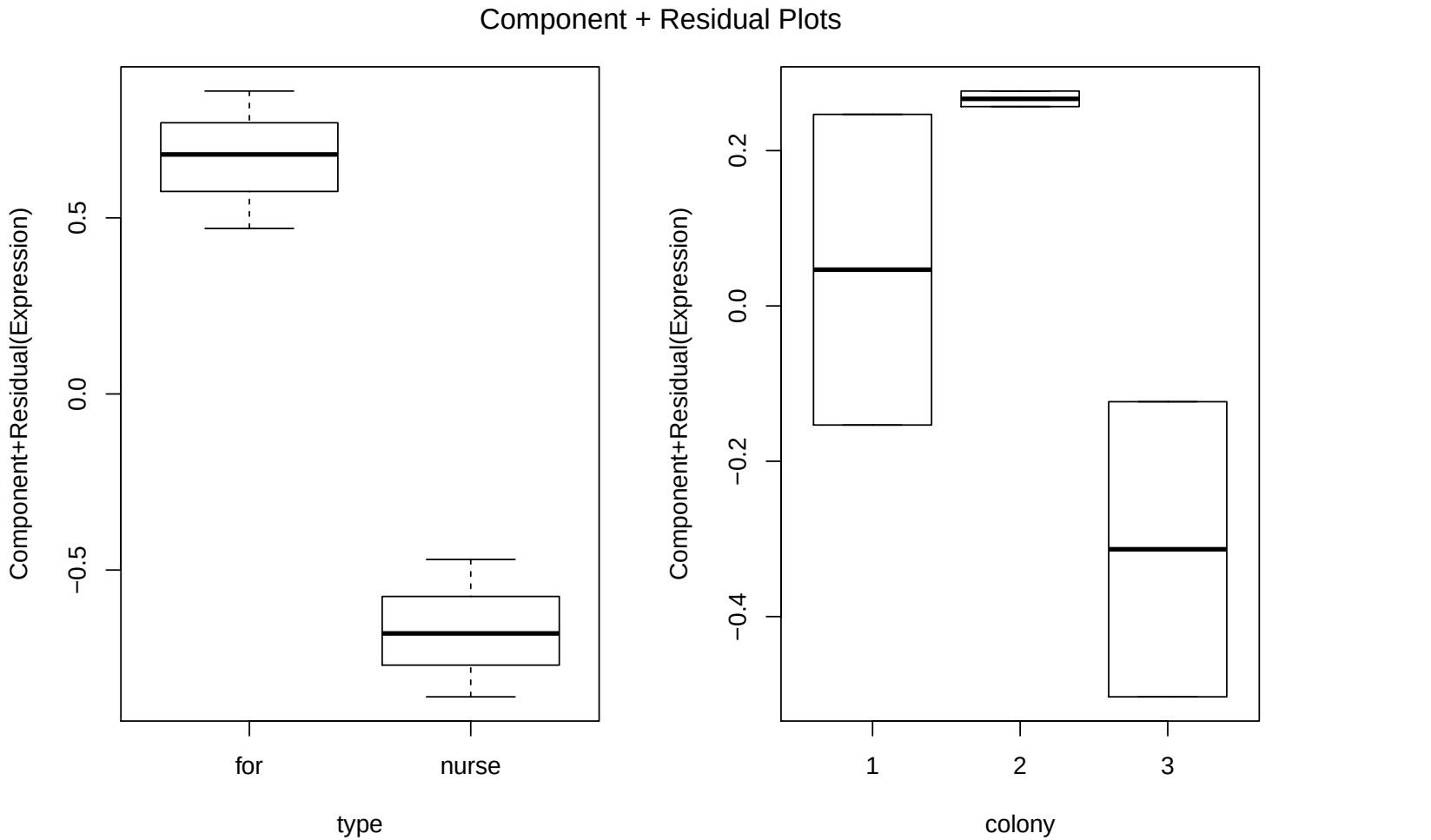
Bee ANOVA

```
anova(bee_lm)

# Analysis of Variance Table
#
# Response: Expression
#           Df Sum Sq Mean Sq F value Pr(>F)
# type       1  2.693   2.693   35.35  0.027
# colony     2  0.343   0.171    2.25  0.308
# Residuals  2  0.152   0.076
```

Bee Effects

```
crPlots(bee_lm)
```



What if my data is unbalanced?

Unbalancing the Zooplankton Data

```
zoop_u <- zoop[-c(1,2),]
```

An Unbalanced ANOVA

```
zoop_u_lm <- update(zoop_lm, data=zoop_u)
anova(zoop_u_lm)

# Analysis of Variance Table
#
# Response: zooplankton
#           Df Sum Sq Mean Sq F value Pr(>F)
# treatment  2   4.18   2.088   16.48 0.0037
# block      4   1.75   0.437    3.45 0.0860
# Residuals  6   0.76   0.127
```

An Unbalanced ANOVA

```
zoop_u_lm <- update(zoop_lm, data=zoop_u)
anova(zoop_u_lm)

# Analysis of Variance Table
#
# Response: zooplankton
#           Df Sum Sq Mean Sq F value Pr(>F)
# treatment  2   4.18   2.088   16.48 0.0037
# block      4   1.75   0.437    3.45 0.0860
# Residuals  6   0.76   0.127
```

Is this valid? Can we use Type I sequential SS?

Unbalanced Data and Type I SS

Missing "cells" (i.e., treatment-block combinations) mean that order matters in testing SS

```
zoop_u_lm1 <- lm(zooplankton ~ treatment + block, data=zoop_u)
zoop_u_lm2 <- lm(zooplankton ~ block + treatment, data=zoop_u)
```

Intercept versus Treatment and Block versus Treatment + Block will not produce different SS

Unbalanced Data and Type I SS

```
# Analysis of Variance Table
#
# Response: zooplankton
#           Df Sum Sq Mean Sq F value Pr(>F)
# treatment  2   4.18   2.088   16.48 0.0037
# block      4   1.75   0.437    3.45 0.0860
# Residuals  6   0.76   0.127

# Analysis of Variance Table
#
# Response: zooplankton
#           Df Sum Sq Mean Sq F value Pr(>F)
# block      4   2.24   0.559    4.41  0.053
# treatment  2   3.69   1.843   14.55  0.005
# Residuals  6   0.76   0.127
```

Solution: Marginal, or Type II SS

SS of Block: Treatment versus Treatment + Block
SS of Treatment: Block versus Block + Treatment

Note: Because of marginality, the sum of all SS will no longer equal SST

Solution: Marginal, or Type II SS

```
Anova(zoop_u_lm1)

# Anova Table (Type II tests)
#
# Response: zooplankton
#           Sum Sq Df F value Pr(>F)
# treatment   3.69  2   14.55  0.005
# block       1.75  4    3.45  0.086
# Residuals   0.76  6
```

Note the capital "A" - this is a function from the car package.