LinearRegression

Temitope Folorunso
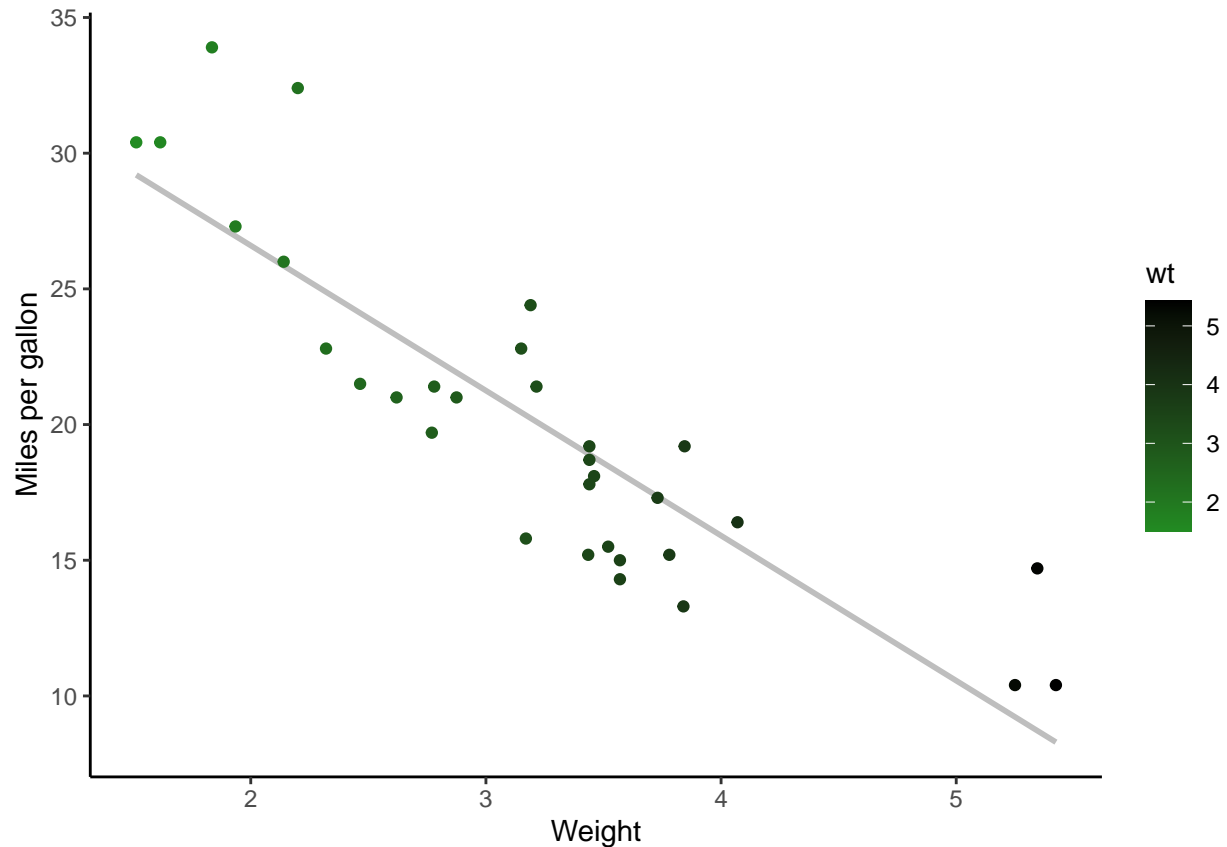
2025-04-01

Contents

Linear Models

Introduction to Regression Analysis in R

```r
data("mtcars")

ggplot(mtcars, aes(x = wt, y = mpg)) +
  geom_smooth(method = lm, se = FALSE, color = "grey") +
  geom_point(aes(color = wt)) +
  xlab("Weight") +
  ylab("Miles per gallon") +
  scale_colour_gradient(low = "forestgreen", high = "black") +
  theme_classic()
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

```
lm1 <- lm(mpg~wt, data = mtcars)
summary(lm1)
```

```
##
## Call:
## lm(formula = mpg ~ wt, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.5432 -2.3647 -0.1252  1.4096  6.8727
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  37.2851     1.8776  19.858  < 2e-16 ***
## wt           -5.3445     0.5591  -9.559 1.29e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.046 on 30 degrees of freedom
## Multiple R-squared:  0.7528, Adjusted R-squared:  0.7446
## F-statistic: 91.38 on 1 and 30 DF,  p-value: 1.294e-10
```

```
anova(lm1)
```

```
## Analysis of Variance Table
```

```
## 
## Response: mpg
##           Df Sum Sq Mean Sq F value    Pr(>F)
## wt         1 847.73  847.73  91.375 1.294e-10 ***
## Residuals 30 278.32    9.28
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
cor.test(mtcars$wt, mtcars$mpg)
```

```
## 
##  Pearson's product-moment correlation
## 
## data:  mtcars$wt and mtcars$mpg
## t = -9.559, df = 30, p-value = 1.294e-10
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.9338264 -0.7440872
## sample estimates:
##        cor 
## -0.8676594
```
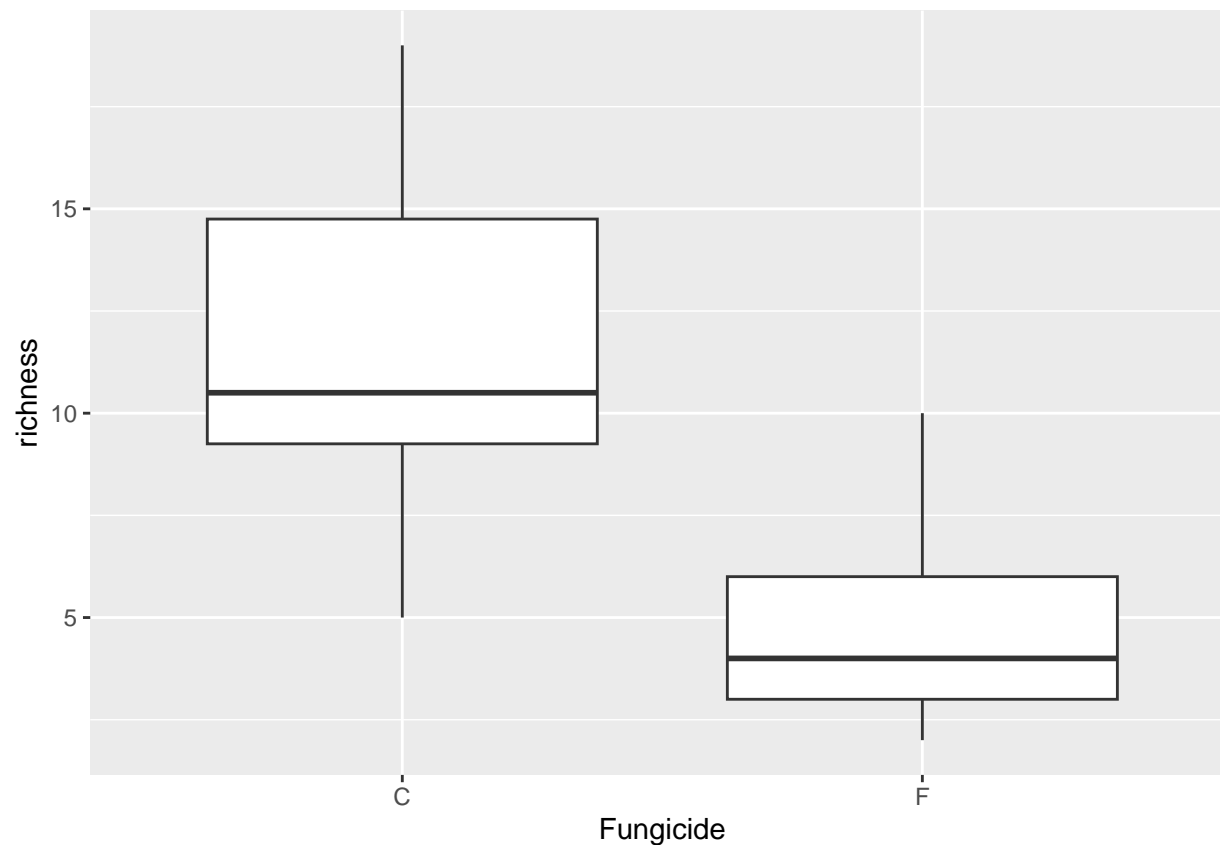
Assumption

```
model <- lm(mpg~wt, data = mtcars)

ggplot(model, aes(y = .resid, x = .fitted)) +
  geom_point() +
  geom_hline(yintercept = 0)
```
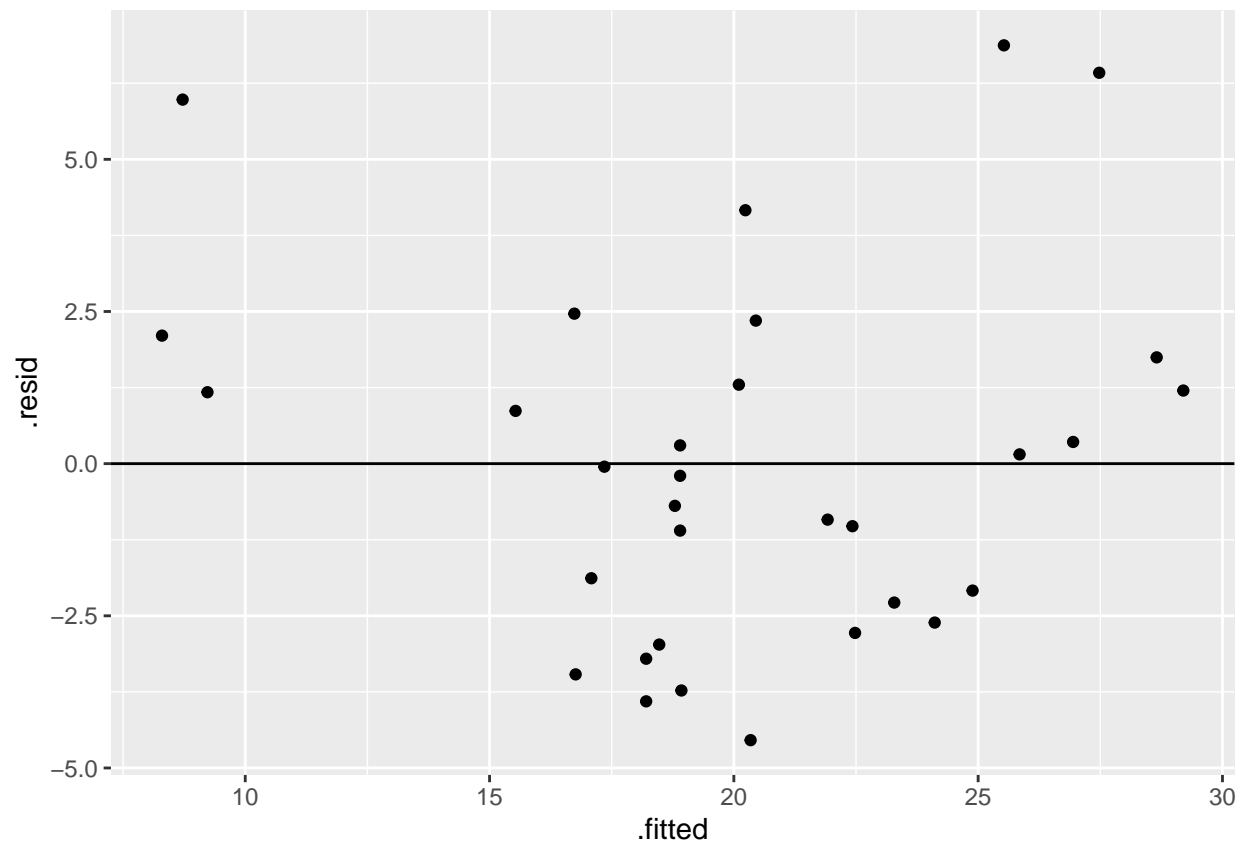
Categorical Variable

```
bull.rich <- read.csv("Data_Folder/Bull_richness.csv")
bull.rich %>%
  filter(GrowthStage == "V8" & Treatment == "Conv.") %>%
  ggplot(aes(x = Fungicide, y = richness)) +
  geom_boxplot()
```

```
bull.rich.subset <- bull.rich %>%
  filter(GrowthStage == "V8" & Treatment == "Conv.")

t.test(richness~ Fungicide, data =bull.rich.subset, var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  richness by Fungicide
## t = 4.8759, df = 22, p-value = 7.118e-05
## alternative hypothesis: true difference in means between group C and group F is not equal to 0
## 95 percent confidence interval:
##   4.118459 10.214875
## sample estimates:
## mean in group C mean in group F
##       11.750000        4.583333
```

```
summary(lm(richness~Fungicide, data =bull.rich.subset))
```

```
##
## Call:
## lm(formula = richness ~ Fungicide, data = bull.rich.subset)
##
## Residuals:
```

```
##      Min      1Q  Median      3Q     Max
## -6.7500 -1.7500 -0.6667  2.2500  7.2500
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   11.750      1.039  11.306 1.24e-10 ***
## FungicideF    -7.167      1.470  -4.876 7.12e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.6 on 22 degrees of freedom
## Multiple R-squared:  0.5194, Adjusted R-squared:  0.4975
## F-statistic: 23.77 on 1 and 22 DF,  p-value: 7.118e-05
```

```r
anova(lm(richness~Fungicide, data = bull.rich.subset))
```

```
## Analysis of Variance Table
##
## Response: richness
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Fungicide  1 308.17 308.167  23.774 7.118e-05 ***
## Residuals 22 285.17  12.962
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

ANOVA

```r
bull.rich.sub2 <- bull.rich %>%
  filter(Fungicide == "C" & Treatment == "Conv." & Crop == "Corn")
bull.rich.sub2$GrowthStage <- factor(bull.rich.sub2$GrowthStage, levels = c("V6", "V8", "V15"))

ggplot(bull.rich.sub2, aes(x = GrowthStage, y = richness)) +
  geom_boxplot()
```
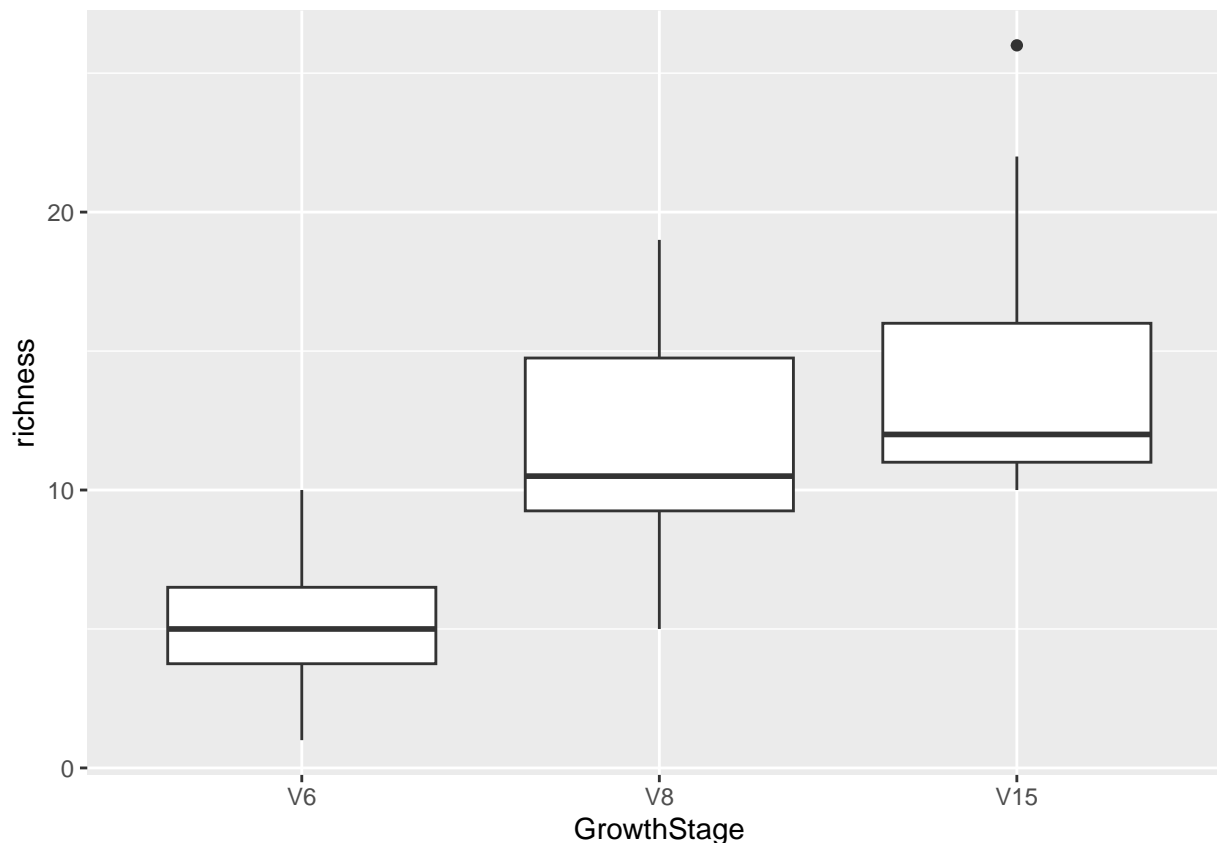
```
lm.growth <- lm(richness ~ GrowthStage, data = bull.rich.sub2)
summary(lm.growth)
```

```
##
## Call:
## lm(formula = richness ~ GrowthStage, data = bull.rich.sub2)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -6.750 -2.625 -1.000  2.250 11.583
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)     5.250      1.208   4.348 0.000124 ***
## GrowthStageV8   6.500      1.708   3.806 0.000581 ***
## GrowthStageV15  9.167      1.708   5.368 6.23e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.183 on 33 degrees of freedom
## Multiple R-squared:  0.4803, Adjusted R-squared:  0.4488
## F-statistic: 15.25 on 2 and 33 DF,  p-value: 2.044e-05
```

```
lmmeans <- emmeans(lm.growth,~GrowthStage)
lmmeans
```

```
## GrowthStage emmean    SE df lower.CL upper.CL
## V6               5.25 1.21 33     2.79     7.71
## V8              11.75 1.21 33     9.29    14.21
## V15             14.42 1.21 33    11.96    16.87
##
## Confidence level used: 0.95
```

```
results_lmeeans <- cld(lmmeans, alpha = 0.05, details = TRUE)
results_lmeeans
```

```
## $emmeans
## GrowthStage emmean    SE df lower.CL upper.CL .group
## V6               5.25 1.21 33     2.79     7.71  1
## V8              11.75 1.21 33     9.29    14.21   2
## V15             14.42 1.21 33    11.96    16.87   2
##
## Confidence level used: 0.95
## P value adjustment: tukey method for comparing a family of 3 estimates
## significance level used: alpha = 0.05
## NOTE: If two or more means share the same grouping symbol,
##       then we cannot show them to be different.
##       But we also did not show them to be the same.
##
## $comparisons
##  contrast estimate    SE df t.ratio p.value
## V8 - V6      6.50 1.71 33   3.806  0.0016
## V15 - V6     9.17 1.71 33   5.368  <.0001
## V15 - V8     2.67 1.71 33   1.562  0.2763
##
## P value adjustment: tukey method for comparing a family of 3 estimates
```

```
anova(lm.growth)
```

```
## Analysis of Variance Table
##
## Response: richness
##             Df Sum Sq Mean Sq F value    Pr(>F)
## GrowthStage  2 533.56 266.778  15.247 2.044e-05 ***
## Residuals   33 577.42  17.497
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(aov(richness ~ GrowthStage, data = bull.rich.sub2))
```

```
##             Df Sum Sq Mean Sq F value  Pr(>F)
## GrowthStage  2  533.6   266.8   15.25 2.04e-05 ***
## Residuals   33  577.4    17.5
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
bull.rich.subset3 <- bull.rich %>%
  filter(Treatment == "Conv." & Crop == "Corn")

lm.inter <- lm(richness ~ GrowthStage + Fungicide + GrowthStage:Fungicide, data = bull.rich.subset3)

# or like this
lm(richness ~ GrowthStage*Fungicide, data = bull.rich.subset3)
```

```
##
## Call:
## lm(formula = richness ~ GrowthStage * Fungicide, data = bull.rich.subset3)
##
## Coefficients:
##            (Intercept)              GrowthStageV6              GrowthStageV8
##                14.4167                    -9.1667                    -2.6667
##              FungicideF  GrowthStageV6:FungicideF  GrowthStageV8:FungicideF
##                -0.9167                    -0.3333                    -6.2500
```

Interactions

```
bull.rich.subset3 <- bull.rich %>%
  filter(Treatment == "Conv." & Crop == "Corn")

lm.inter <- lm(richness ~ GrowthStage + Fungicide + GrowthStage:Fungicide, data = bull.rich.subset3)

# or like this
lm(richness ~ GrowthStage*Fungicide, data = bull.rich.subset3)
```

```
##
## Call:
## lm(formula = richness ~ GrowthStage * Fungicide, data = bull.rich.subset3)
##
## Coefficients:
##            (Intercept)              GrowthStageV6              GrowthStageV8
##                14.4167                    -9.1667                    -2.6667
##              FungicideF  GrowthStageV6:FungicideF  GrowthStageV8:FungicideF
##                -0.9167                    -0.3333                    -6.2500
```

```
summary(lm.inter) # significant terms
```

```
##
## Call:
## lm(formula = richness ~ GrowthStage + Fungicide + GrowthStage:Fungicide,
##     data = bull.rich.subset3)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.5000 -2.4167 -0.4167  2.0625 11.5833
##
## Coefficients:
```

```
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)               14.4167     1.1029  13.072  < 2e-16 ***
## GrowthStageV6             -9.1667     1.5597  -5.877 1.51e-07 ***
## GrowthStageV8             -2.6667     1.5597  -1.710   0.0920 .
## FungicideF                -0.9167     1.5597  -0.588   0.5587
## GrowthStageV6:FungicideF  -0.3333     2.2057  -0.151   0.8803
## GrowthStageV8:FungicideF  -6.2500     2.2057  -2.834   0.0061 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.82 on 66 degrees of freedom
## Multiple R-squared:  0.5903, Adjusted R-squared:  0.5593
## F-statistic: 19.02 on 5 and 66 DF,  p-value: 1.144e-11
```

```r
anova(lm.inter) # The interaction term is signifant.
```
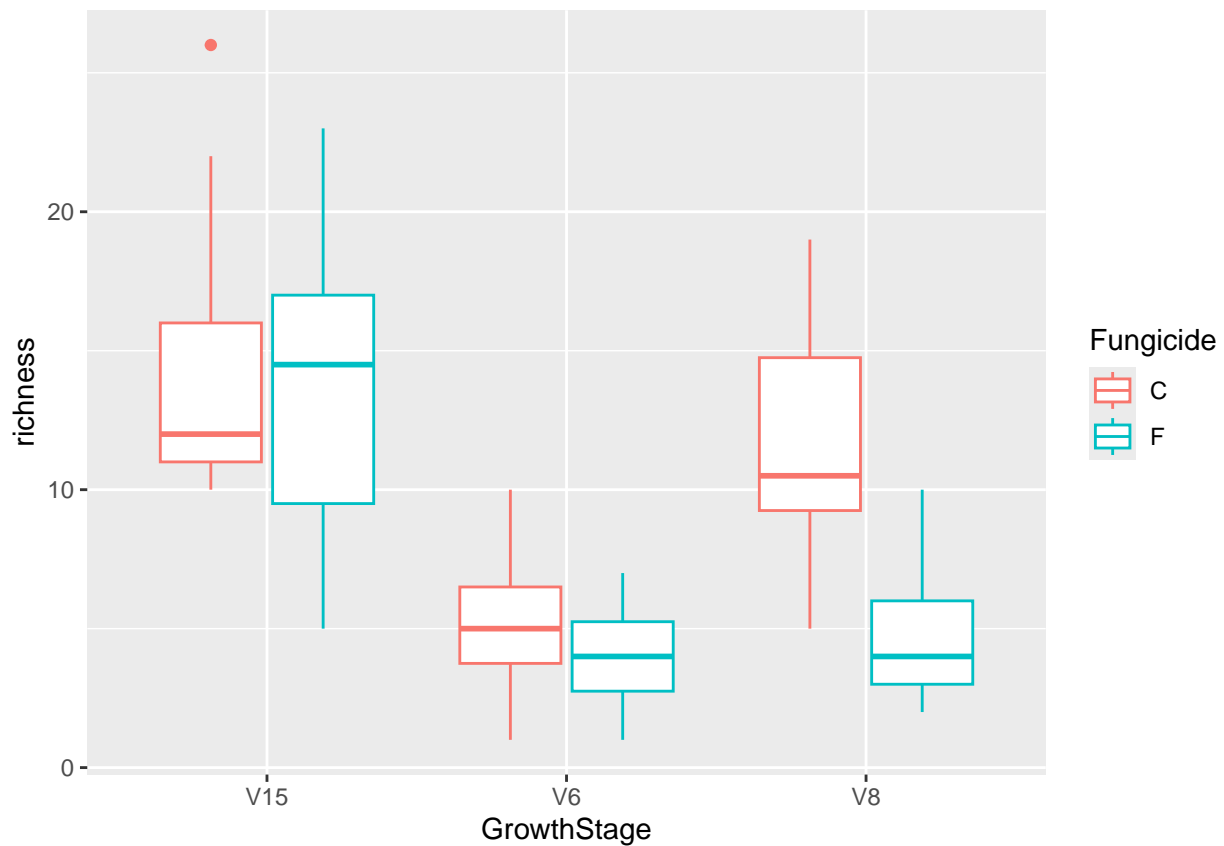
```
## Analysis of Variance Table
##
## Response: richness
##                     Df  Sum Sq Mean Sq F value     Pr(>F)
## GrowthStage          2 1065.58  532.79 36.5027 2.113e-11 ***
## Fungicide            1  174.22  174.22 11.9363 0.0009668 ***
## GrowthStage:Fungicide 2  148.36   74.18  5.0823 0.0088534 **
## Residuals           66  963.33   14.60
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
lsmeans <- emmeans(lm.inter, ~Fungicide|GrowthStage) # estimate lsmeans of variety within siteXyear
Results_lsmeans <- cld(lsmeans, alpha = 0.05, reversed = TRUE, details = TRUE) # contrast with Tukey aj
Results_lsmeans
```

```
## $emmeans
## GrowthStage = V15:
##  Fungicide emmean  SE df lower.CL upper.CL .group
##  C          14.42 1.1 66    12.21    16.62  1
##  F          13.50 1.1 66    11.30    15.70  1
##
## GrowthStage = V6:
##  Fungicide emmean  SE df lower.CL upper.CL .group
##  C           5.25 1.1 66     3.05     7.45  1
##  F           4.00 1.1 66     1.80     6.20  1
##
## GrowthStage = V8:
##  Fungicide emmean  SE df lower.CL upper.CL .group
##  C          11.75 1.1 66     9.55    13.95  1
##  F           4.58 1.1 66     2.38     6.79   2
##
## Confidence level used: 0.95
## significance level used: alpha = 0.05
## NOTE: If two or more means share the same grouping symbol,
##       then we cannot show them to be different.
##       But we also did not show them to be the same.
##
```

```
## $comparisons
## GrowthStage = V15:
##  contrast estimate   SE df t.ratio p.value
##  C - F        0.917 1.56 66   0.588  0.5587
##
## GrowthStage = V6:
##  contrast estimate   SE df t.ratio p.value
##  C - F        1.250 1.56 66   0.801  0.4258
##
## GrowthStage = V8:
##  contrast estimate   SE df t.ratio p.value
##  C - F        7.167 1.56 66   4.595  <.0001
```

```
bull.rich.subset3 %>%
  ggplot(aes(x = GrowthStage, y = richness, color = Fungicide)) +
  geom_boxplot()
```


Mixed effects models