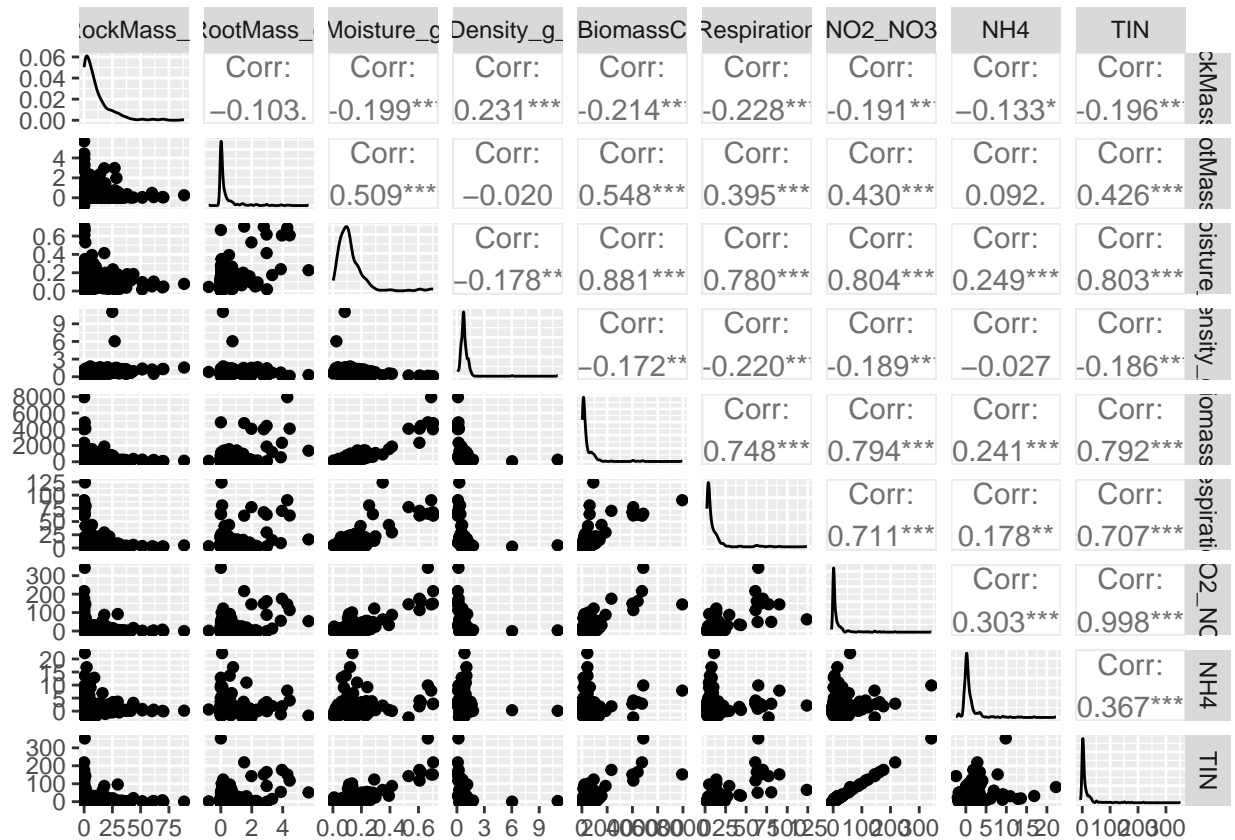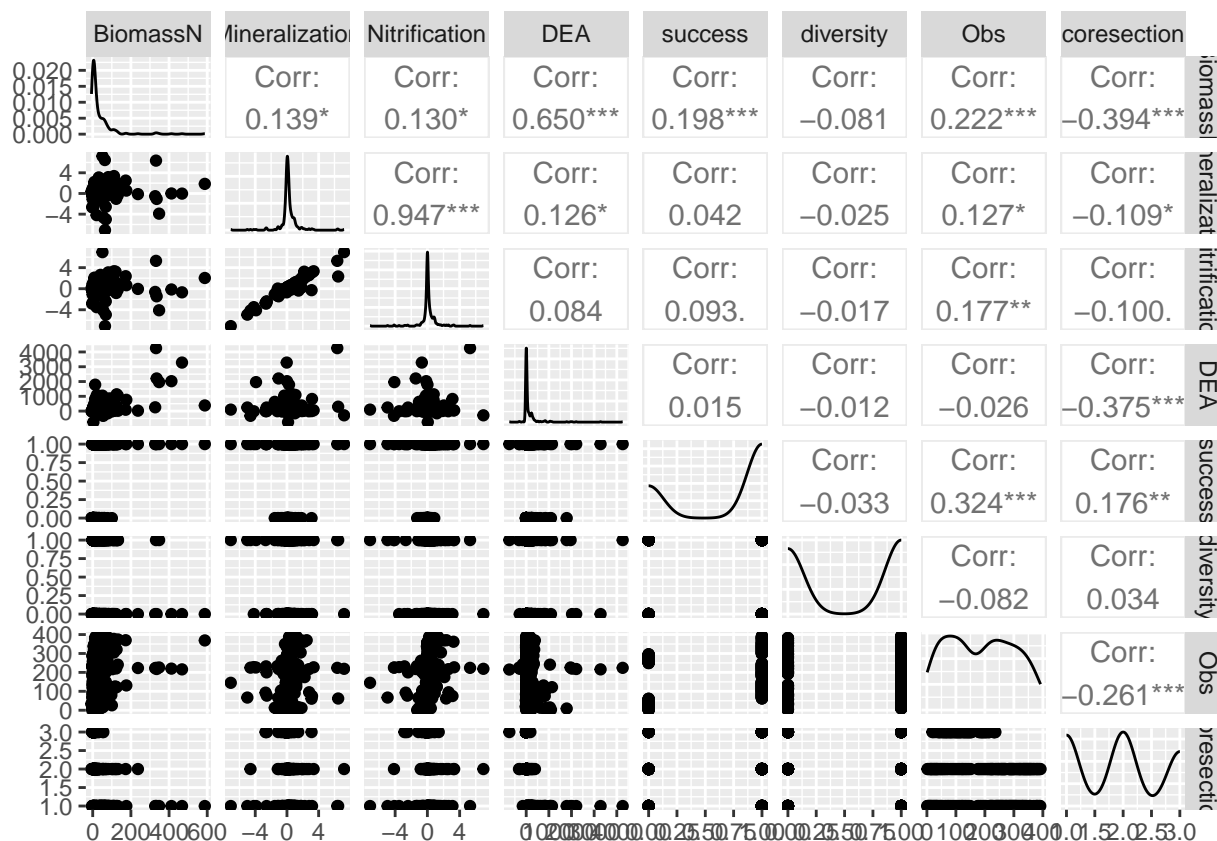```r
tree <- read.csv("C:\\Users\\Siyu Sun\\Documents\\ENV710_Project\\Data\\MTNYCData_modified2_noDEAna.csv
# tree <- read.csv("MTNYCData_modified2_noDEAna.csv")
pacman::p_load(dplyr, ggplot2, lme4, lmerTest, lmtest, GGally, car, DHARMa, boot,
               viridis, ggpubr, performance, MuMIn, merTools, lattice)

# Data exploration
ggpairs(tree[,c(9:17)])
```



```r
ggpairs(tree[,c(18:25)])
```

```r
# Model 1: Biomass C
## Full model
lm1 <- lmer(log(BiomassC) ~ factor(Success) + factor(Diversity) + factor(Season) +
            log(Respiration) + log(BiomassN + 6.53271102 + 0.001) +
            log(NO2_NO3 + 5.9671545 + 0.001) + log(NH4 + 2.422020695 + 0.001) +
            log(RootMass_g + 0.782 + 0.001) + log(Moisture_g) + factor(coresection) +
            factor(Year) + (1|Site/Plot), data = tree)
summary(lm1)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula:
## log(BiomassC) ~ factor(Success) + factor(Diversity) + factor(Season) +
##     log(Respiration) + log(BiomassN + 6.53271102 + 0.001) + log(NO2_NO3 +
##     5.9671545 + 0.001) + log(NH4 + 2.422020695 + 0.001) + log(RootMass_g +
##     0.782 + 0.001) + log(Moisture_g) + factor(coresection) +
##     factor(Year) + (1 | Site/Plot)
##    Data: tree
##
## REML criterion at convergence: 358.6
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.1699 -0.5652 -0.0227  0.5614  3.4270
##
## Random effects:
```

```
##  Groups    Name        Variance Std.Dev.
##  Plot:Site (Intercept) 0.005058 0.07112
##  Site      (Intercept) 0.024876 0.15772
##  Residual              0.143130 0.37833
## Number of obs: 326, groups:  Plot:Site, 36; Site, 11
##
## Fixed effects:
##                                     Estimate Std. Error          df t value
## (Intercept)                         5.392253    0.294327 215.081868  18.321
## factor(Success)Low                 -0.122796    0.126842   9.345961  -0.968
## factor(Diversity)Low                0.122035    0.049385  38.427402   2.471
## factor(Season)Summer               -0.147818    0.143457 219.384690  -1.030
## log(Respiration)                    0.453756    0.058551 283.892480   7.750
## log(BiomassN + 6.53271102 + 0.001)  0.142611    0.028656 311.903280   4.977
## log(NO2_NO3 + 5.9671545 + 0.001)    0.024933    0.032321 306.507401   0.771
## log(NH4 + 2.422020695 + 0.001)     -0.042003    0.035469 311.837518  -1.184
## log(RootMass_g + 0.782 + 0.001)    -0.007843    0.044102 309.089692  -0.178
## log(Moisture_g)                     0.384181    0.048379 309.628276   7.941
## factor(coresection)2               -0.316397    0.070281 296.037765  -4.502
## factor(coresection)3               -0.547229    0.095111 300.405679  -5.754
## factor(Year)2019                    0.055468    0.136603 244.509502   0.406
##                                    Pr(>|t|)
## (Intercept)                        < 2e-16 ***
## factor(Success)Low                   0.357
## factor(Diversity)Low                 0.018 *
## factor(Season)Summer                 0.304
## log(Respiration)                   1.65e-13 ***
## log(BiomassN + 6.53271102 + 0.001) 1.07e-06 ***
## log(NO2_NO3 + 5.9671545 + 0.001)     0.441
## log(NH4 + 2.422020695 + 0.001)       0.237
## log(RootMass_g + 0.782 + 0.001)      0.859
## log(Moisture_g)                    3.73e-14 ***
## factor(coresection)2               9.70e-06 ***
## factor(coresection)3               2.15e-08 ***
## factor(Year)2019                     0.685
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## Model reduction
lm2 <- update(lm1, ~.-log(RootMass_g + 0.782 + 0.001))
# summary(lm2)
lm3 <- update(lm2, ~.-factor(Year))
# summary(lm3)
lm4 <- update(lm3, ~.-log(NO2_NO3 + 5.9671545 + 0.001))
# summary(lm4)
lm5 <- update(lm4, ~.-factor(Success))
# summary(lm5)
lm6 <- update(lm5, ~.-log(NH4 + 2.422020695 + 0.001))
summary(lm6)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula:
## log(BiomassC) ~ factor(Diversity) + factor(Season) + log(Respiration) +
```

```
##      log(BiomassN + 6.53271102 + 0.001) + log(Moisture_g) + factor(coresection) +
##      (1 | Site/Plot)
##   Data: tree
##
## REML criterion at convergence: 342.9
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.1089 -0.5721 -0.0160  0.5705  3.4651
##
## Random effects:
##  Groups     Name        Variance Std.Dev.
##  Plot:Site (Intercept) 0.005305 0.07284
##  Site      (Intercept) 0.028005 0.16735
##  Residual              0.141670 0.37639
## Number of obs: 326, groups:  Plot:Site, 36; Site, 11
##
## Fixed effects:
##                                  Estimate Std. Error        df t value
## (Intercept)                       5.41920    0.22625 241.42824  23.952
## factor(Diversity)Low              0.11538    0.04899  37.99347   2.355
## factor(Season)Summer             -0.20218    0.05874 305.86191  -3.442
## log(Respiration)                  0.46672    0.05651 287.73048   8.260
## log(BiomassN + 6.53271102 + 0.001)  0.14222    0.02798 314.76318   5.083
## log(Moisture_g)                   0.38984    0.04682 316.76033   8.327
## factor(coresection)2             -0.31111    0.06933 298.09918  -4.487
## factor(coresection)3             -0.53608    0.09201 283.34052  -5.826
##                                  Pr(>|t|)
## (Intercept)                       < 2e-16 ***
## factor(Diversity)Low             0.023778 *
## factor(Season)Summer             0.000659 ***
## log(Respiration)                 5.38e-15 ***
## log(BiomassN + 6.53271102 + 0.001) 6.39e-07 ***
## log(Moisture_g)                  2.53e-15 ***
## factor(coresection)2             1.03e-05 ***
## factor(coresection)3             1.54e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##            (Intr) fc(D)L fc(S)S lg(Rs) l(+6+0 lg(M_) fct()2
## fctr(Dvrs)L -0.158
## fctr(Ssn)Sm  0.021  0.106
## log(Rsprtn) -0.595 -0.100 -0.415
## l(BN+6.53+0 -0.500  0.137  0.218 -0.220
## log(Mstr_g)  0.728 -0.087  0.065 -0.327 -0.280
## fctr(crsc)2 -0.490 -0.035 -0.180  0.559  0.153 -0.005
## fctr(crsc)3 -0.396 -0.052 -0.434  0.558  0.153  0.087  0.718
```

```r
### delete the outlier according to the outlier plot
tree1 <- tree[-c(33),]
lm7 <- lmer(log(BiomassC) ~ factor(Diversity) + factor(Season) + log(Respiration) +
            log(BiomassN + 4.79350047 + 0.001) + log(Moisture_g) + factor(coresection) +
            (1 | Site/Plot), data = tree1)
```

```
summary(lm7)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula:
## log(BiomassC) ~ factor(Diversity) + factor(Season) + log(Respiration) +
##     log(BiomassN + 4.79350047 + 0.001) + log(Moisture_g) + factor(coresection) +
##     (1 | Site/Plot)
##    Data: tree1
##
## REML criterion at convergence: 337.1
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.1815 -0.5497 -0.0251  0.5824  3.1173
##
## Random effects:
##  Groups     Name        Variance Std.Dev.
##  Plot:Site  (Intercept) 0.00500  0.07071
##  Site       (Intercept) 0.02702  0.16439
##  Residual               0.13964  0.37369
## Number of obs: 325, groups:  Plot:Site, 36; Site, 11
##
## Fixed effects:
##                                   Estimate Std. Error       df t value
## (Intercept)                        5.41408    0.21967 238.62915  24.646
## factor(Diversity)Low               0.11787    0.04852  39.35465   2.429
## factor(Season)Summer              -0.19314    0.05844 305.64278  -3.305
## log(Respiration)                   0.45474    0.05624 286.97207   8.085
## log(BiomassN + 4.79350047 + 0.001) 0.14891    0.02618 314.72491   5.688
## log(Moisture_g)                    0.38299    0.04644 315.42414   8.246
## factor(coresection)2              -0.31234    0.06875 297.44784  -4.543
## factor(coresection)3              -0.52582    0.09131 281.45830  -5.758
##                                   Pr(>|t|)
## (Intercept)                        < 2e-16 ***
## factor(Diversity)Low               0.01979 *
## factor(Season)Summer               0.00106 **
## log(Respiration)                   1.76e-14 ***
## log(BiomassN + 4.79350047 + 0.001) 2.94e-08 ***
## log(Moisture_g)                    4.46e-15 ***
## factor(coresection)2               8.07e-06 ***
## factor(coresection)3               2.22e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##            (Intr) fc(D)L fc(S)S lg(Rs) l(+4+0 lg(M_) fct()2
## fctr(Dvrs)L -0.157
## fctr(Ssn)Sm  0.026  0.110
## log(Rsprtn) -0.609 -0.100 -0.416
## l(BN+4.79+0 -0.465  0.137  0.225 -0.235
## log(Mstr_g)  0.732 -0.088  0.063 -0.323 -0.280
## fctr(crsc)2 -0.488 -0.033 -0.179  0.559  0.136 -0.002
```

```
## fctr(crsc)3 -0.399 -0.051 -0.430  0.551  0.157  0.085  0.716
```

```
AIC(lm1, lm2, lm3,lm4,lm5,lm6, lm7)
```

```
##     df      AIC
## lm1 16 390.6419
## lm2 15 384.2674
## lm3 14 380.2646
## lm4 13 373.7761
## lm5 12 370.5006
## lm6 11 364.9400
## lm7 11 359.0947
```

```
## Assumptions
### dept vars normality
shapiro.test(tree$BiomassC)
```

```
##
##   Shapiro-Wilk normality test
##
## data:  tree$BiomassC
## W = 0.44773, p-value < 2.2e-16
```

```
shapiro.test(log(tree$BiomassC))
```

```
##
##   Shapiro-Wilk normality test
##
## data:  log(tree$BiomassC)
## W = 0.94443, p-value = 9.979e-10
```

```
### Residuals mean & normality
mean(residuals(lm6))
```
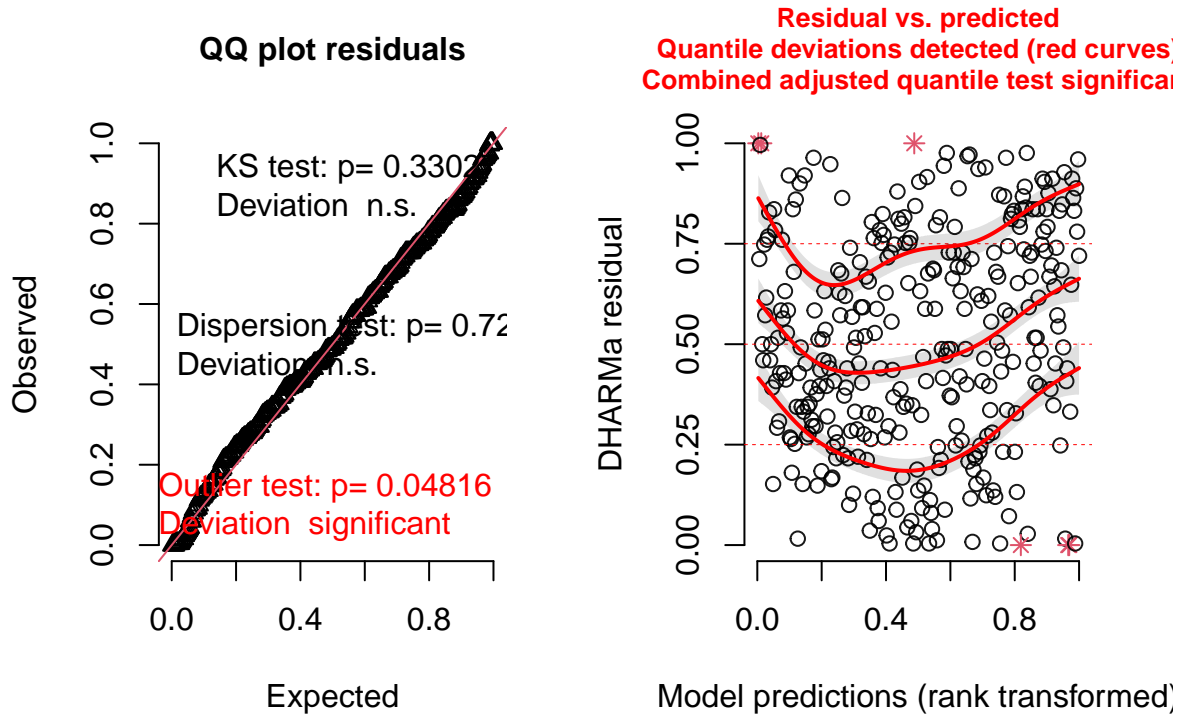
```
## [1] 3.0582e-15
```

```
shapiro.test(residuals(lm6))
```

```
##
##   Shapiro-Wilk normality test
##
## data:  residuals(lm6)
## W = 0.99099, p-value = 0.04346
```

```
lmm6 <- simulateResiduals(fittedModel = lm6)
plot(lmm6)
```

DHARMa residual

QQ plot residuals

KS test: p= 0.3302
Deviation n.s.

Dispersion test: p= 0.72
Deviation n.s.

Outlier test: p= 0.04816
Deviation significant

Observed

Expected

Residual vs. predicted
Quantile deviations detected (red curves)
Combined adjusted quantile test significar

DHARMa residual

Model predictions (rank transformed)

```
check_model(lm6, check = c("reqq", "qq", "linearity"))
```

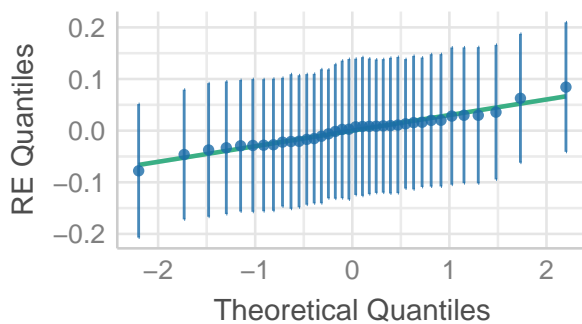## Linearity
Reference line should be flat and horizontal
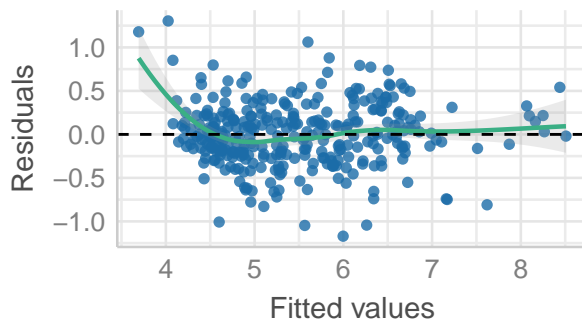
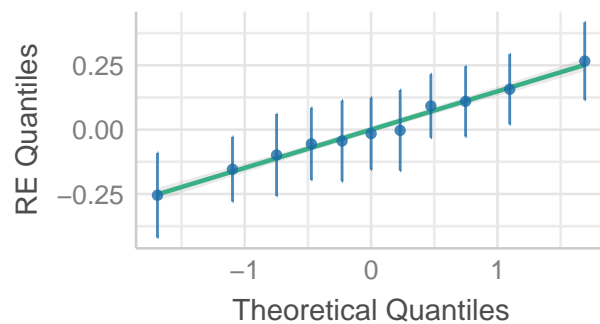

Normality of Residuals
Dots should fall along the line
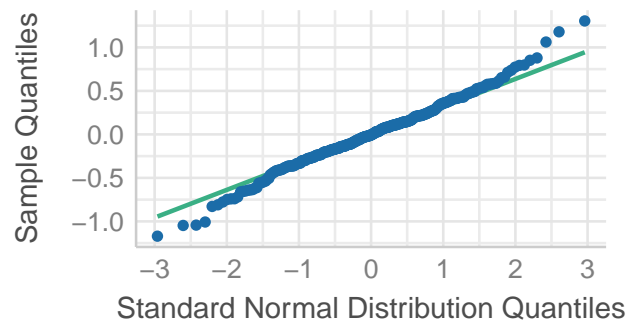
Normality of Random Effects (Plot:Site)
Dots should be plotted along the line

Normality of Random Effects (Site)
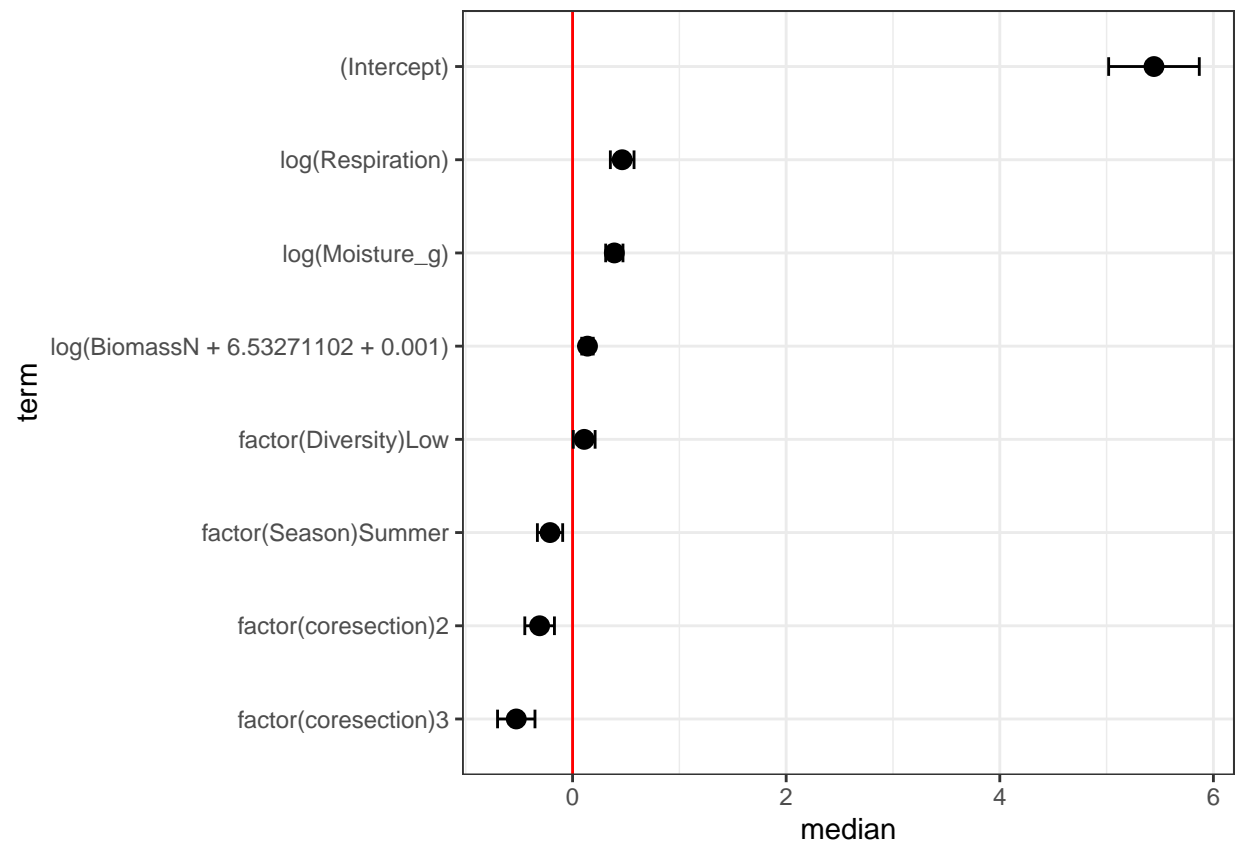Dots should be plotted along the line


```
### Goodness of fit / R2 FE RE
r.squaredGLMM(lm6)
```

```
##              R2m        R2c
## [1,] 0.8158701 0.8509218
```

```
plotFEsim(FEsim(lm6), intercept = T)
```
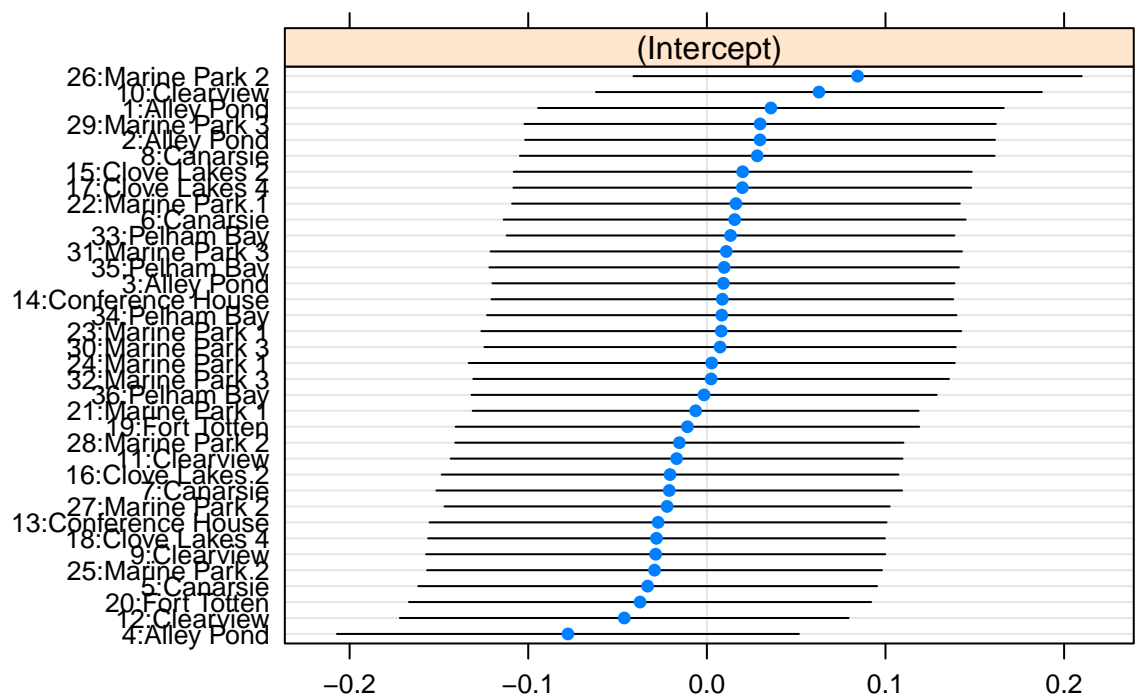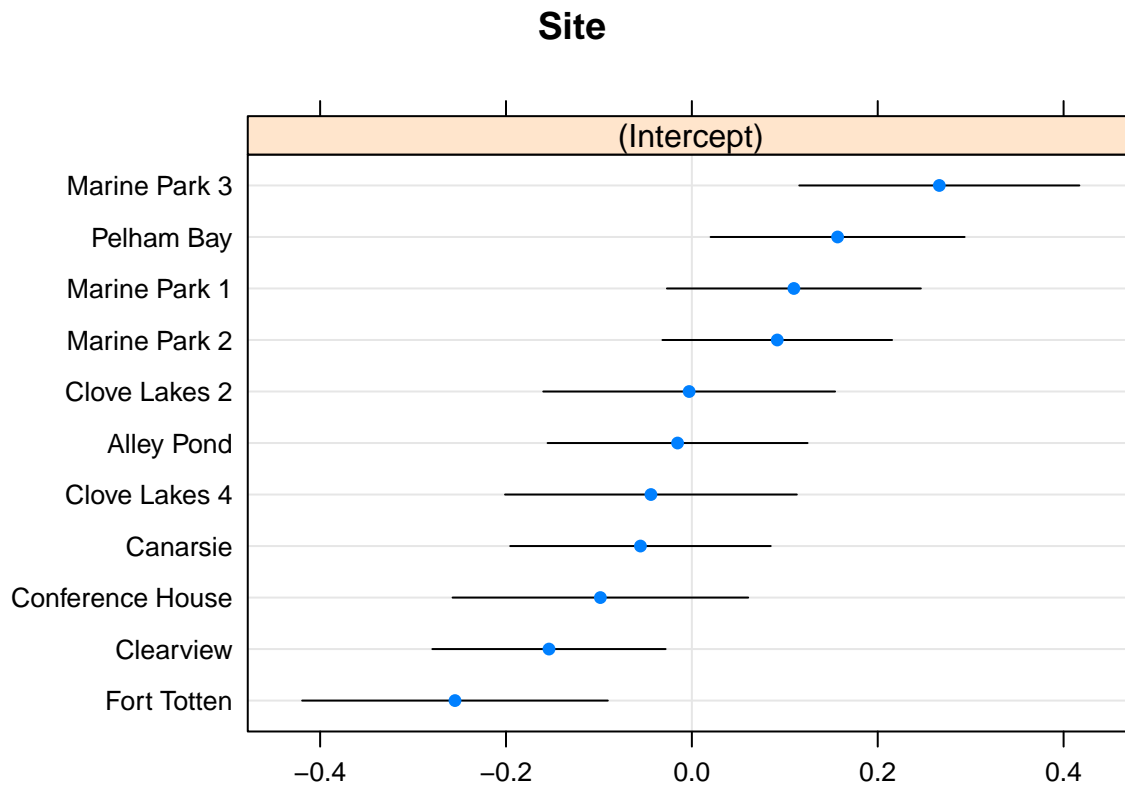
```
dotplot(ranef(lm6))
```
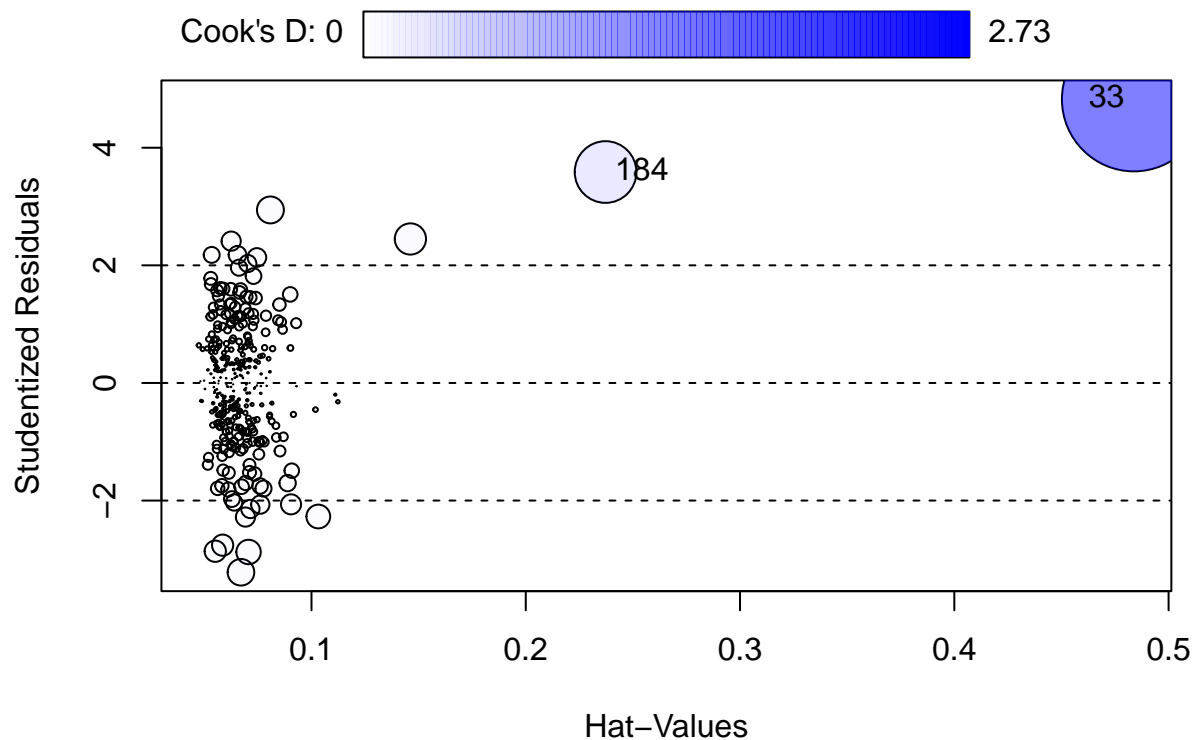
```
## $`Plot:Site`
```

Plot:Site


```
## 
## $Site
```

Site

Outlier test

```
outlierTest(lm6)
```

```
##    rstudent unadjusted p-value Bonferroni p
## 33 4.823018       2.2083e-06   0.00071991
```

```
influencePlot(lm6)
```

```
##       StudRes       Hat      CookD
## 33   4.823018 0.4838408 2.7256277
## 184 3.587067 0.2372410 0.5002549
```

Multicollinearity
```
vif(lm6)
```
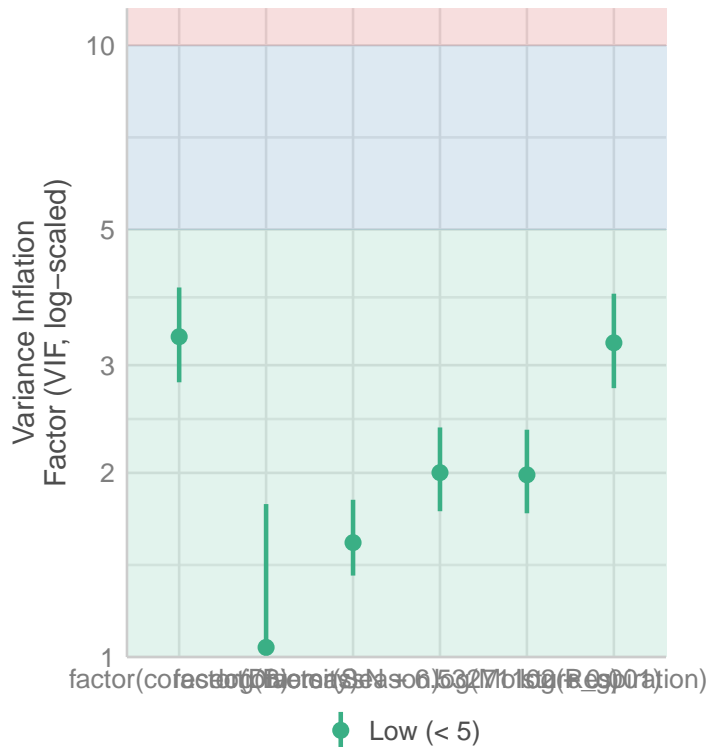
```
##                                         GVIF Df GVIF^(1/(2*Df))
## factor(Diversity)                    1.037072  1        1.018367
## factor(Season)                       1.538391  1        1.240319
## log(Respiration)                     3.263823  1        1.806605
## log(BiomassN + 6.53271102 + 0.001)   2.002028  1        1.414930
## log(Moisture_g)                      1.985524  1        1.409086
## factor(coresection)                  3.339383  2        1.351813
```

```
check_model(lm6, check = c("vif"))
```

Collinearity
High collinearity (VIF) may inflate parameter uncertainty

factor(cofactor(fol)ctea:ts(aN os6,)532l71t:llog(lRe_gp0ra)tion)

Low (< 5)

Graphs

Respiration-diversity
```
a <- fixef(lm7)

eq.re <- function(x){exp(a[1]+a[2]+a[4]*log(x)+ a[5]*mean(log(tree1$BiomassN +
                                               4.79350047 + 0.001))+
                    a[6]*mean(log(tree1$Moisture_g)))}

eq.re1 <- function(x){exp(a[1]+a[4]*log(x)+ a[5]*mean(log(tree1$BiomassN +
                                               4.79350047 + 0.001))+
                    a[6]*mean(log(tree1$Moisture_g)))}

library(viridis)
ggplot(tree1, aes(x=Respiration, y=BiomassC)) +
  geom_jitter(height = 0.02, width = 0.02, alpha = 0.3, size = 3) +
  stat_function(fun=eq.re, geom="line", linewidth=0.7, aes(color="Low")) +
  stat_function(fun=eq.re1, geom="line", linewidth=0.7, aes(color="High")) +
  theme_bw() +
  theme(legend.position = "right") +
  labs(y = "BiomassC", x = "Respiration") +
  scale_color_viridis("Diversity", discrete = TRUE)
```
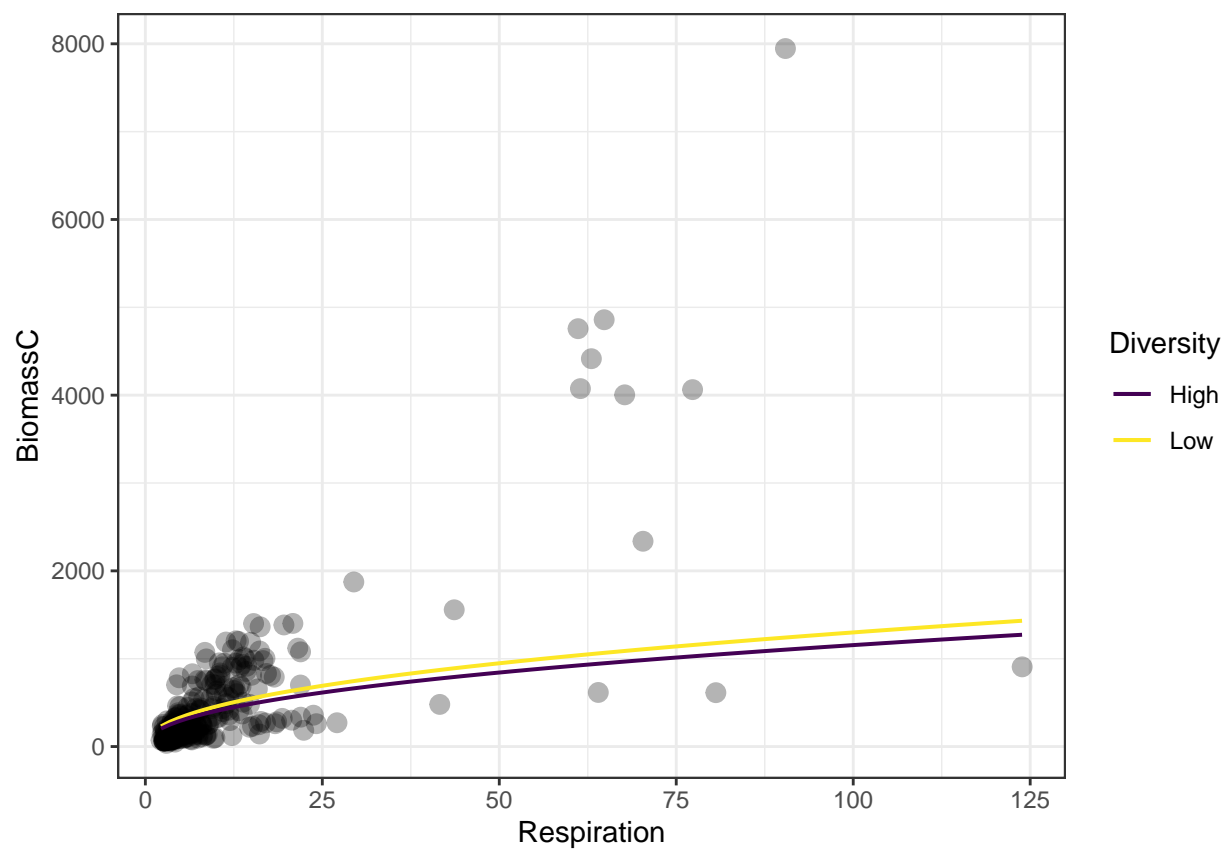
Figure 1: Effects of Respiration and Diversity on Biomass C

```
## Moisture-diversity
a <- fixef(lm7)

eq.mo <- function(x){exp(a[1]+a[2]+a[4]*mean(log(tree1$Respiration))+
                         a[5]*mean(log(tree1$BiomassN + 4.79350047 + 0.001))+
                         a[6]*log(x))}

eq.mo1 <- function(x){exp(a[1]+a[4]*mean(log(tree1$Respiration))+
                          a[5]*mean(log(tree1$BiomassN + 4.79350047 + 0.001))+
                          a[6]*log(x))}

ggplot(tree1, aes(x=Moisture_g, y=BiomassC)) +
  geom_jitter(height = 0.02, width = 0.02, alpha = 0.3, size = 3) +
  stat_function(fun=eq.mo, geom="line", linewidth=0.7, aes(color="Low")) +
  stat_function(fun=eq.mo1, geom="line", linewidth=0.7, aes(color="High")) +
  theme_bw() +
  theme(legend.position = "right") +
  labs(y = "BiomassC", x = "Moisture") +
  scale_color_viridis("Diversity", discrete = TRUE)
```
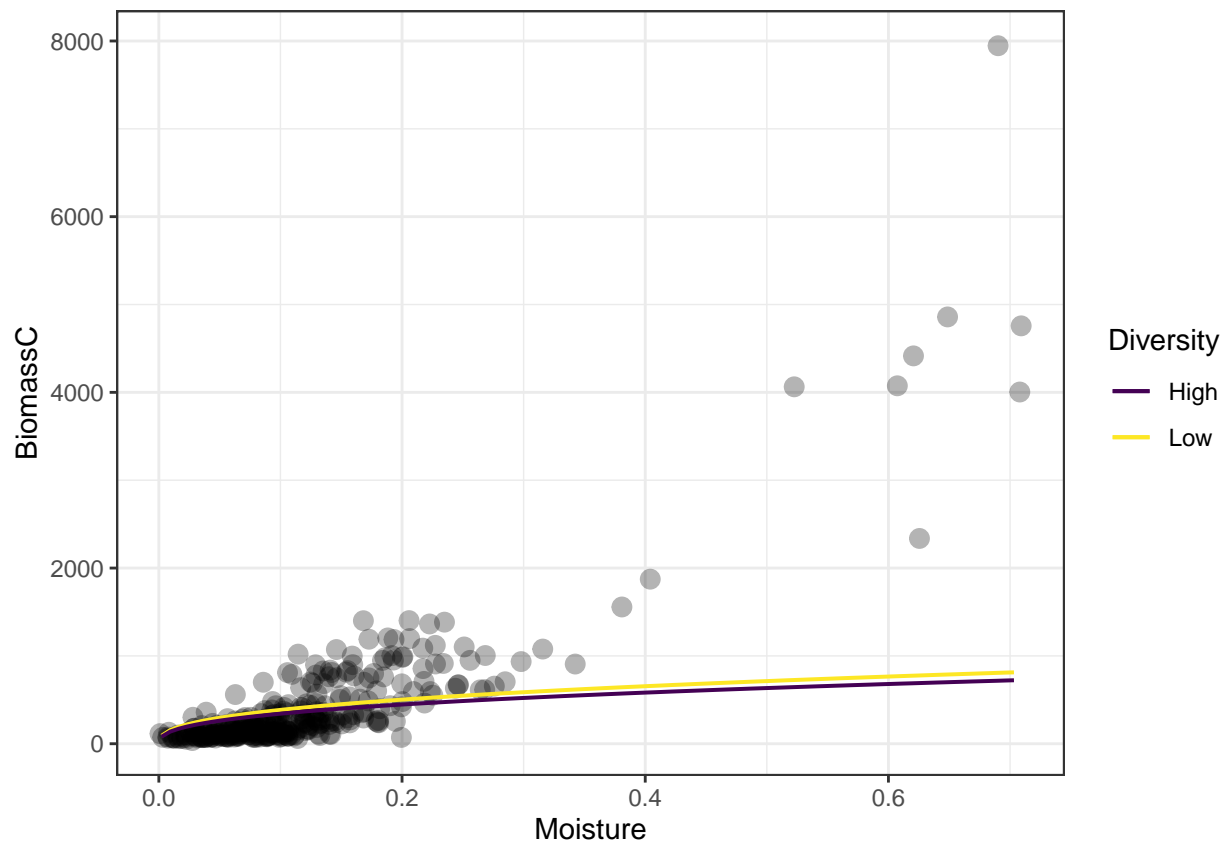

Figure 2: Effects of Moisture and Diversity on Biomass C

```
# Calculate mean and standard error for each year
mean_data <- aggregate(tree1$BiomassC, by=list(group=tree1$coresection), FUN=mean)
```

```
names(mean_data)[2] <- "Mean"
se_data <- aggregate(tree1$BiomassC, by=list(group=tree1$coresection), FUN=sd)
se_data$se <- se_data$x / sqrt(length(tree1$BiomassC))

# Create bar plot with error bars
ggplot(mean_data, aes(x=factor(group), y=Mean, fill=factor(group))) +
  geom_bar(stat="identity", position="dodge", alpha = 0.5) +
  geom_errorbar(aes(ymin=Mean-se_data$se, ymax=Mean+se_data$se),width=.2) +
  labs(x="Core Section Groups", y="Biomass",fill = "Core Section") +
  scale_fill_discrete(labels=c("0-10cm","10-30cm","above 30cm")) +
  theme_bw()
```
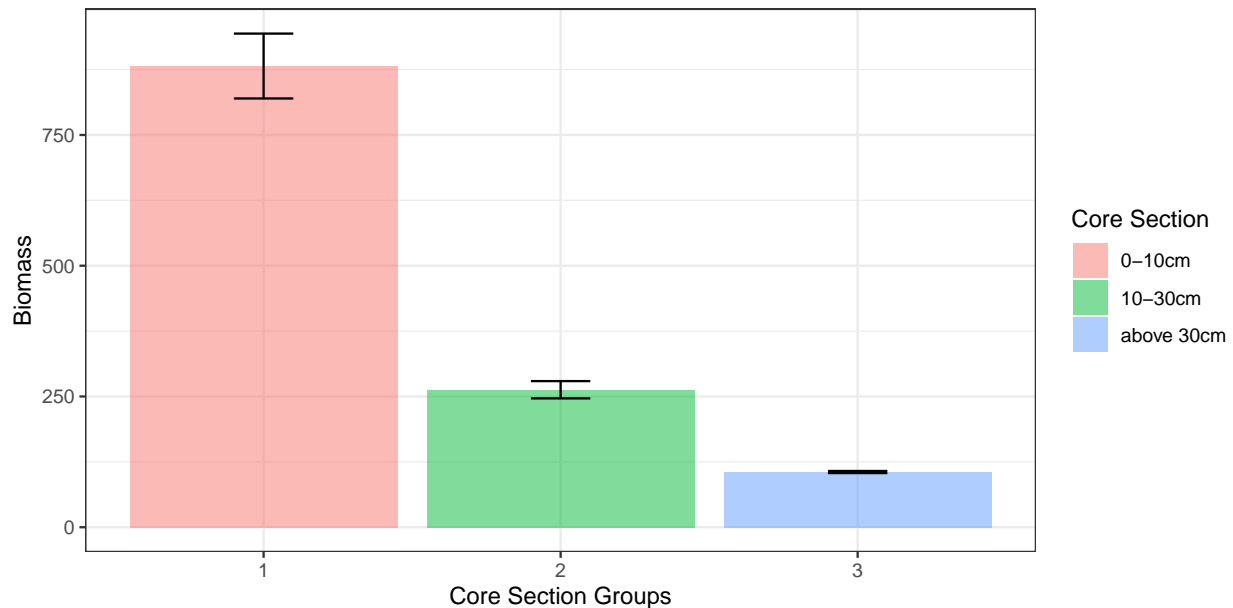

Figure 3: Mean biomass carbon and differences among core section groups

```
# Model 2: Biomass N
## log-transform all of the variables
tree.l <- tree
for (i in 1:326){
  tree.l$RockMass_g[i] <- log(tree$RockMass_g[i] + abs(min(tree$RockMass_g)) + 0.001)
  tree.l$RootMass_g[i] <- log(tree$RootMass_g[i] + abs(min(tree$RootMass_g)) + 0.001)
  tree.l$Moisture_g[i] <- log(tree$Moisture_g[i] + abs(min(tree$Moisture_g)) + 0.001)
  tree.l$BulkDensity_g_cm3[i] <- log(tree$BulkDensity_g_cm3[i] +
                                     abs(min(tree$BulkDensity_g_cm3)) + 0.001)
  tree.l$BiomassC[i] <- log(tree$BiomassC[i] + abs(min(tree$BiomassC)) + 0.001)
  tree.l$Respiration[i] <- log(tree$Respiration[i] + abs(min(tree$Respiration)) + 0.001)
  tree.l$NO2_NO3[i] <- log(tree$NO2_NO3[i] + abs(min(tree$NO2_NO3)) + 0.001)
  tree.l$NH4[i] <- log(tree$NH4[i] + abs(min(tree$NH4)) + 0.001)
  tree.l$TIN[i] <- log(tree$TIN[i] + abs(min(tree$TIN)) + 0.001)
  tree.l$BiomassN[i] <- log(tree$BiomassN[i] + abs(min(tree$BiomassN)) + 0.001)
  tree.l$DEA[i] <- log(tree$DEA[i] + abs(min(tree$DEA)) + 0.001)
}
```

```
### Remove highly influencial observations
no.out.tree.l <- tree.l[-c(33,29,193),]

## Full model
noY.finalm1 <- lmer(BiomassN ~ Success + factor(diversity) + factor(Season) +
                      Respiration + NO2_NO3 + NH4 + Nitrification + RootMass_g +
                      Moisture_g + factor(coresection) + BiomassC +
                      (1|Site/Plot), data = no.out.tree.l)
summary(noY.finalm1)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula:
## BiomassN ~ Success + factor(diversity) + factor(Season) + Respiration +
##     NO2_NO3 + NH4 + Nitrification + RootMass_g + Moisture_g +
##     factor(coresection) + BiomassC + (1 | Site/Plot)
##    Data: no.out.tree.l
##
## REML criterion at convergence: 501.5
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.2929 -0.4786 -0.1267  0.3839  4.5732
##
## Random effects:
##  Groups     Name        Variance Std.Dev.
##  Plot:Site (Intercept) 0.00884  0.09402
##  Site      (Intercept) 0.01408  0.11865
##  Residual              0.23336  0.48307
## Number of obs: 323, groups:  Plot:Site, 36; Site, 11
##
## Fixed effects:
##                       Estimate Std. Error        df t value Pr(>|t|)
## (Intercept)           -1.27923    0.56884 242.22678  -2.249 0.025422 *
## SuccessLow            -0.09383    0.11631   9.25990  -0.807 0.440015
## factor(diversity)1     0.19304    0.06323  33.86557   3.053 0.004390 **
## factor(Season)Summer  -0.20107    0.07620 283.52603  -2.639 0.008788 **
## Respiration           -0.01427    0.09812 203.94676  -0.145 0.884509
## NO2_NO3                0.36367    0.06000 182.17324   6.062 7.58e-09 ***
## NH4                    0.21553    0.06811 305.32924   3.165 0.001709 **
## Nitrification          0.10537    0.02985 309.75867   3.531 0.000478 ***
## RootMass_g             0.07047    0.05575 306.97349   1.264 0.207170
## Moisture_g             0.06011    0.07362 267.09297   0.816 0.414996
## factor(coresection)2  -0.09706    0.09010 260.57695  -1.077 0.282369
## factor(coresection)3  -0.09549    0.12280 245.31825  -0.778 0.437588
## BiomassC               0.62116    0.07839 296.76521   7.924 4.68e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## Model reduction
noY.finalm2 <- update(noY.finalm1,.~.-Respiration)
summary(noY.finalm2)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: BiomassN ~ Success + factor(diversity) + factor(Season) + NO2_NO3 +
##     NH4 + Nitrification + RootMass_g + Moisture_g + factor(coresection) +
##     BiomassC + (1 | Site/Plot)
##    Data: no.out.tree.l
##
## REML criterion at convergence: 498.7
##
## Scaled residuals:
##     Min     1Q  Median     3Q     Max
## -3.3021 -0.4869 -0.1233  0.3865  4.5734
##
## Random effects:
##  Groups     Name        Variance Std.Dev.
##  Plot:Site (Intercept) 0.00899  0.09482
##  Site      (Intercept) 0.01344  0.11593
##  Residual              0.23268  0.48236
## Number of obs: 323, groups:  Plot:Site, 36; Site, 11
##
## Fixed effects:
##                        Estimate Std. Error        df t value Pr(>|t|)
## (Intercept)            -1.28772    0.56721 243.18907  -2.270 0.024067 *
## SuccessLow             -0.09422    0.11493   9.32558  -0.820 0.432809
## factor(diversity)1      0.19341    0.06318  33.91497   3.061 0.004290 **
## factor(Season)Summer   -0.20453    0.07067 293.85237  -2.894 0.004085 **
## NO2_NO3                 0.36155    0.05705 173.68509   6.337 1.95e-09 ***
## NH4                     0.21451    0.06793 306.42970   3.158 0.001747 **
## Nitrification           0.10491    0.02954 310.91302   3.551 0.000443 ***
## RootMass_g              0.07073    0.05566 307.95813   1.271 0.204767
## Moisture_g              0.05892    0.07330 264.79420   0.804 0.422285
## factor(coresection)2   -0.09196    0.08443 279.58877  -1.089 0.277032
## factor(coresection)3   -0.09024    0.11830 267.35798  -0.763 0.446232
## BiomassC                0.61751    0.07264 297.19521   8.501 9.29e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) SccssL fct()1 fc(S)S NO2_NO NH4    Ntrfct RtMss_ Mstr_g
## SuccessLow  -0.173
## fctr(dvrs)1 -0.141 -0.003
## fctr(Ssn)Sm -0.095 -0.140 -0.022
## NO2_NO3     -0.093  0.266 -0.076 -0.230
## NH4         -0.141 -0.084  0.011  0.098 -0.153
## Nitrificatn -0.018  0.093 -0.017  0.063  0.254  0.139
## RootMass_g   0.000 -0.056  0.063  0.152 -0.074  0.076  0.036
## Moisture_g   0.733  0.075  0.009 -0.051 -0.128 -0.190 -0.061 -0.079
## fctr(crsc)2 -0.523  0.205  0.069  0.021  0.045 -0.010 -0.059  0.008 -0.065
## fctr(crsc)3 -0.514  0.331  0.088 -0.268  0.201 -0.042  0.040  0.065 -0.027
## BiomassC    -0.892  0.084  0.131  0.082 -0.252 -0.014 -0.134 -0.045 -0.519
##             fct()2 fct()3
## SuccessLow
## fctr(dvrs)1
## fctr(Ssn)Sm
```

```
## NO2_NO3
## NH4
## Nitrificatn
## RootMass_g
## Moisture_g
## fctr(crsc)2
## fctr(crsc)3  0.686
## BiomassC      0.523  0.500
```

```
noY.finalm3 <- update(noY.finalm2,.~.-Success)
summary(noY.finalm3)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: BiomassN ~ factor(diversity) + factor(Season) + NO2_NO3 + NH4 +
##     Nitrification + RootMass_g + Moisture_g + factor(coresection) +
##     BiomassC + (1 | Site/Plot)
##    Data: no.out.tree.l
##
## REML criterion at convergence: 496.8
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.3494 -0.4690 -0.1112  0.4068  4.5191
##
## Random effects:
##  Groups     Name        Variance Std.Dev.
##  Plot:Site (Intercept) 0.00880  0.09381
##  Site      (Intercept) 0.01217  0.11030
##  Residual              0.23291  0.48260
## Number of obs: 323, groups:  Plot:Site, 36; Site, 11
##
## Fixed effects:
##                        Estimate Std. Error         df t value Pr(>|t|)
## (Intercept)            -1.38395    0.55706 246.21245  -2.484 0.013644 *
## factor(diversity)1      0.19294    0.06304  33.77976   3.061 0.004308 **
## factor(Season)Summer   -0.21137    0.06990 287.17198  -3.024 0.002720 **
## NO2_NO3                 0.37563    0.05471 141.34605   6.866 1.92e-10 ***
## NH4                     0.20881    0.06763 305.98483   3.088 0.002202 **
## Nitrification           0.10762    0.02940 311.82191   3.661 0.000295 ***
## RootMass_g              0.06885    0.05554 309.38033   1.240 0.216038
## Moisture_g              0.06220    0.07291 260.38359   0.853 0.394430
## factor(coresection)2   -0.07583    0.08244 254.67402  -0.920 0.358506
## factor(coresection)3   -0.05595    0.11114 173.41306  -0.503 0.615290
## BiomassC                0.62389    0.07227 295.29895   8.632 3.80e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##            (Intr) fct()1 fc(S)S NO2_NO NH4    Ntrfct RtMss_ Mstr_g fct()2
## fctr(dvrs)1 -0.143
## fctr(Ssn)Sm -0.123 -0.023
## NO2_NO3     -0.044 -0.078 -0.201
## NH4         -0.161  0.010  0.088 -0.135
```

```
## Nitrificatn   0.000 -0.016  0.078  0.237  0.149
## RootMass_g   -0.008  0.063  0.145 -0.064  0.072  0.040
## Moisture_g    0.761  0.009 -0.040 -0.153 -0.185 -0.067 -0.074
## fctr(crsc)2 -0.504  0.072  0.053 -0.017  0.010 -0.082  0.019 -0.082
## fctr(crsc)3 -0.490  0.095 -0.236  0.116 -0.012  0.007  0.088 -0.055  0.668
## BiomassC     -0.895  0.132  0.095 -0.291 -0.005 -0.144 -0.041 -0.529  0.517
##              fct()3
## fctr(dvrs)1
## fctr(Ssn)Sm
## NO2_NO3
## NH4
## Nitrificatn
## RootMass_g
## Moisture_g
## fctr(crsc)2
## fctr(crsc)3
## BiomassC      0.502
```

```
noY.finalm4 <- update(noY.finalm3,.~.-factor(coresection))
summary(noY.finalm4)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: BiomassN ~ factor(diversity) + factor(Season) + NO2_NO3 + NH4 +
##     Nitrification + RootMass_g + Moisture_g + BiomassC + (1 |     Site/Plot)
##    Data: no.out.tree.l
##
## REML criterion at convergence: 491.4
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.4138 -0.4663 -0.1334  0.4184  4.4573
##
## Random effects:
##  Groups     Name        Variance Std.Dev.
##  Plot:Site (Intercept) 0.009034 0.09505
##  Site      (Intercept) 0.009846 0.09923
##  Residual              0.232621 0.48231
## Number of obs: 323, groups:  Plot:Site, 36; Site, 11
##
## Fixed effects:
##                        Estimate Std. Error        df t value Pr(>|t|)
## (Intercept)           -1.64083    0.46539 289.72398  -3.526 0.000491 ***
## factor(diversity)1     0.19575    0.06284  33.73312   3.115 0.003739 **
## factor(Season)Summer  -0.20151    0.06484 301.14516  -3.108 0.002065 **
## NO2_NO3                0.37584    0.05336 160.66118   7.044 5.19e-11 ***
## NH4                    0.20744    0.06745 308.22688   3.075 0.002290 **
## Nitrification          0.10574    0.02914 313.79443   3.629 0.000332 ***
## RootMass_g             0.07005    0.05513 311.31184   1.271 0.204810
## Moisture_g             0.05469    0.07225 254.23902   0.757 0.449756
## BiomassC               0.65706    0.05978 303.93913  10.991  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Correlation of Fixed Effects:
##            (Intr) fct()1 fc(S)S NO2_NO NH4    Ntrfct RtMss_ Mstr_g
## fctr(dvrs)1 -0.113
## fctr(Ssn)Sm -0.219 -0.003
## NO2_NO3     -0.017 -0.088 -0.152
## NH4         -0.195  0.011  0.086 -0.130
## Nitrificatn -0.027 -0.014  0.121  0.221  0.155
## RootMass_g   0.030  0.056  0.195 -0.088  0.076  0.032
## Moisture_g   0.861  0.016 -0.040 -0.156 -0.185 -0.074 -0.070
## BiomassC    -0.850  0.098  0.187 -0.390 -0.006 -0.144 -0.088 -0.589
```

```
noY.finalm5 <- update(noY.finalm4,.~.-Moisture_g)
summary(noY.finalm5)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: BiomassN ~ factor(diversity) + factor(Season) + NO2_NO3 + NH4 +
##     Nitrification + RootMass_g + BiomassC + (1 | Site/Plot)
##    Data: no.out.tree.l
##
## REML criterion at convergence: 488.5
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.3991 -0.4676 -0.1344  0.4386  4.4735
##
## Random effects:
##  Groups     Name        Variance Std.Dev.
##  Plot:Site (Intercept) 0.008849 0.09407
##  Site      (Intercept) 0.008804 0.09383
##  Residual              0.232769 0.48246
## Number of obs: 323, groups:  Plot:Site, 36; Site, 11
##
## Fixed effects:
##                        Estimate Std. Error        df t value Pr(>|t|)
## (Intercept)            -1.94655    0.23636 264.01834  -8.236 8.33e-15 ***
## factor(diversity)1      0.19456    0.06268  33.87752   3.104 0.003842 **
## factor(Season)Summer   -0.19817    0.06474 300.99653  -3.061 0.002404 **
## NO2_NO3                 0.38330    0.05245 159.86995   7.308 1.22e-11 ***
## NH4                     0.21572    0.06621 306.95307   3.258 0.001247 **
## Nitrification           0.10778    0.02904 314.74527   3.712 0.000243 ***
## RootMass_g              0.07344    0.05496 312.92493   1.336 0.182391
## BiomassC                0.68370    0.04822 283.88098  14.178  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##            (Intr) fct()1 fc(S)S NO2_NO NH4    Ntrfct RtMss_
## fctr(dvrs)1 -0.249
## fctr(Ssn)Sm -0.364 -0.002
## NO2_NO3      0.236 -0.085 -0.161
## NH4         -0.072  0.014  0.081 -0.162
## Nitrificatn  0.072 -0.013  0.119  0.210  0.145
## RootMass_g   0.177  0.058  0.192 -0.102  0.065  0.027
```

```
## BiomassC      -0.835   0.132   0.202 -0.605 -0.144 -0.232 -0.160
```

```
FinalModel <- update(noY.finalm5,.~.-RootMass_g)

AIC(noY.finalm1, noY.finalm2, noY.finalm3, noY.finalm4, noY.finalm5, FinalModel)
```

```
##              df      AIC
## noY.finalm1 16 533.4508
## noY.finalm2 15 528.6632
## noY.finalm3 14 524.8332
## noY.finalm4 12 515.3555
## noY.finalm5 11 510.4977
## FinalModel  10 506.3081
```

```
## Assumptions
### Residuals mean & normality
mean(residuals(FinalModel))
```
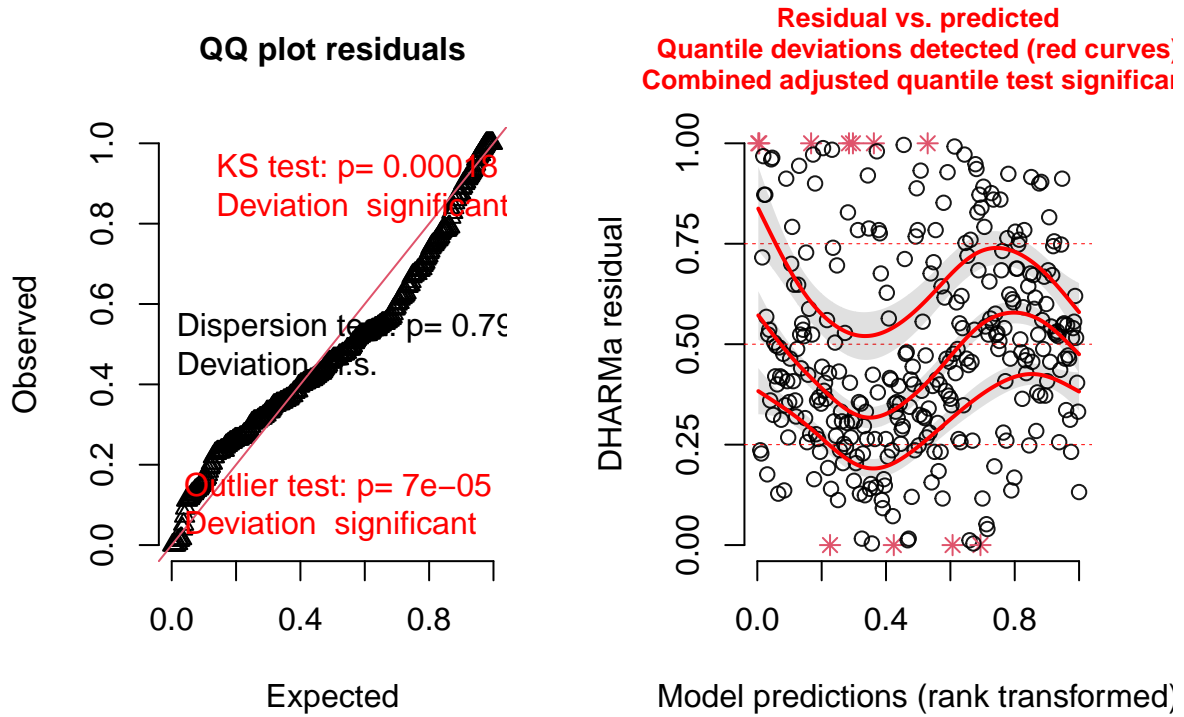
```
## [1] 1.622579e-16
```

```
shapiro.test(residuals(FinalModel))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(FinalModel)
## W = 0.92215, p-value = 6.413e-12
```

```
library(DHARMa)
lmm6 <- simulateResiduals(fittedModel = FinalModel)
plot(lmm6)
```
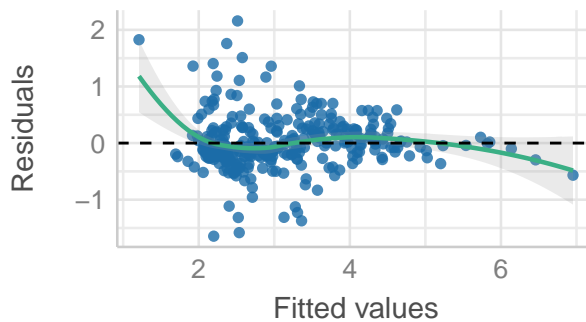
DHARMa residual

QQ plot residuals

KS test: p= 0.00013
Deviation significant

Dispersion test: p= 0.79
Deviation n.s.

Outlier test: p= 7e−05
Deviation significant

Observed

Expected

Residual vs. predicted
Quantile deviations detected (red curves)
Combined adjusted quantile test significant

DHARMa residual

Model predictions (rank transformed)

```
library(performance)
performance::check_model(FinalModel, check = c("reqq", "qq", "linearity"))
```

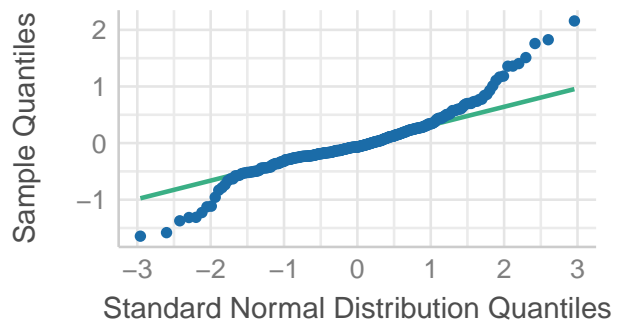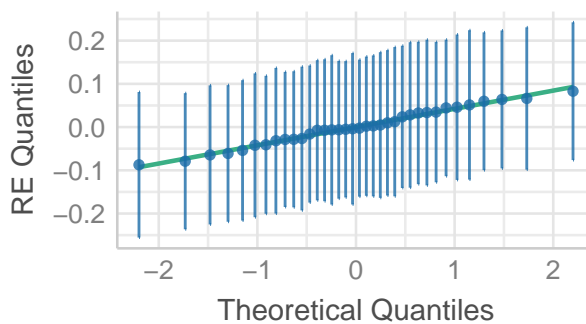## Linearity
Reference line should be flat and horizontal

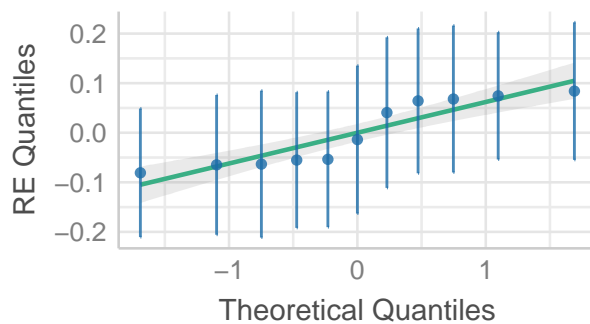

Normality of Residuals
Dots should fall along the line

Normality of Random Effects (Plot:Site)
Dots should be plotted along the line

Normality of Random Effects (Site)
Dots should be plotted along the line
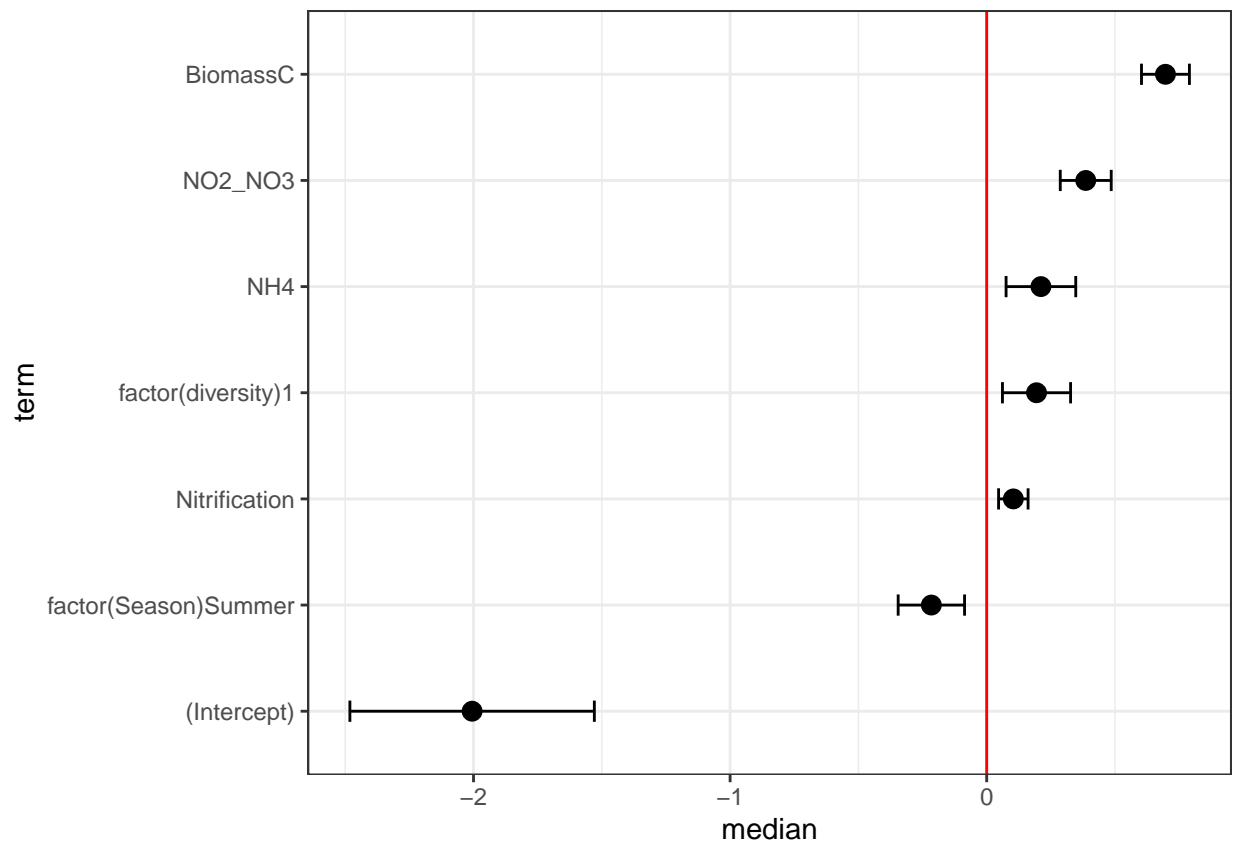


```
### Goodness of fit / R2 FE RE
library(merTools)
merTools::plotFEsim(FEsim(FinalModel), intercept = T)
```
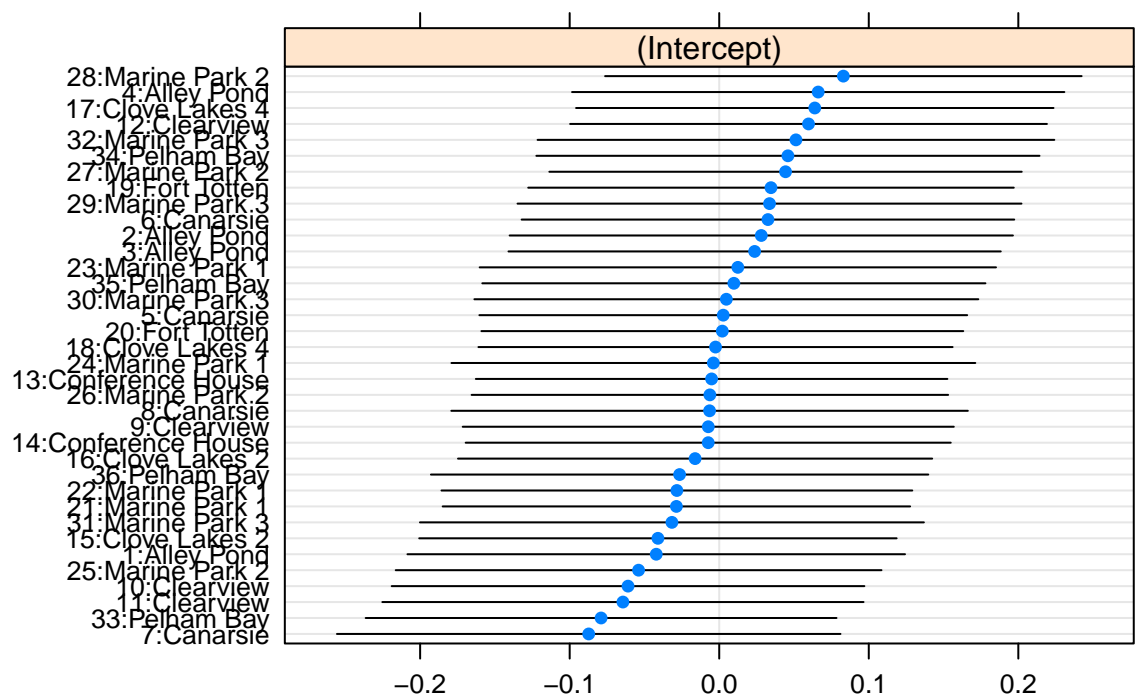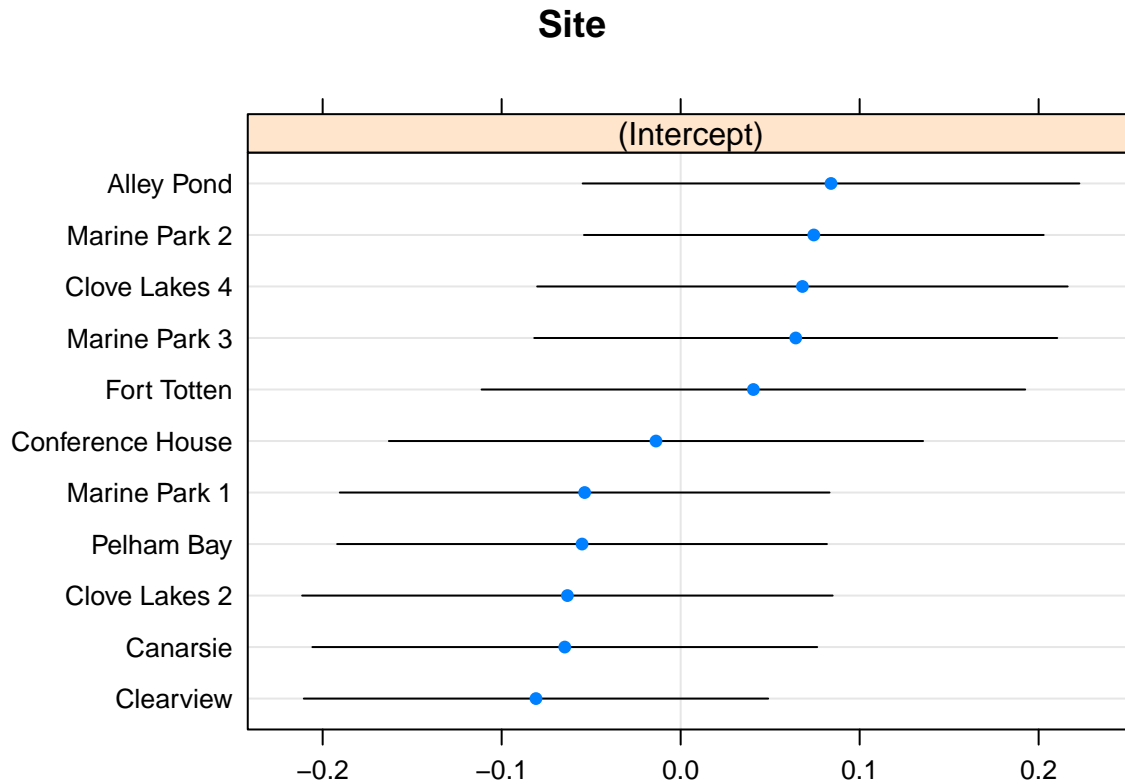
```
library(lattice)
dotplot(ranef(FinalModel))
```

```
## $`Plot:Site`
```

Plot:Site


```
## 
## $Site
```

Site


```
### Outlier test
library(car)
outlierTest(FinalModel)
```

```
##     rstudent unadjusted p-value Bonferroni p
## 24 4.587167         6.5154e-06    0.0021045
## 27 3.973749         8.7937e-05    0.0284040
```

```
influencePlot(FinalModel)
```
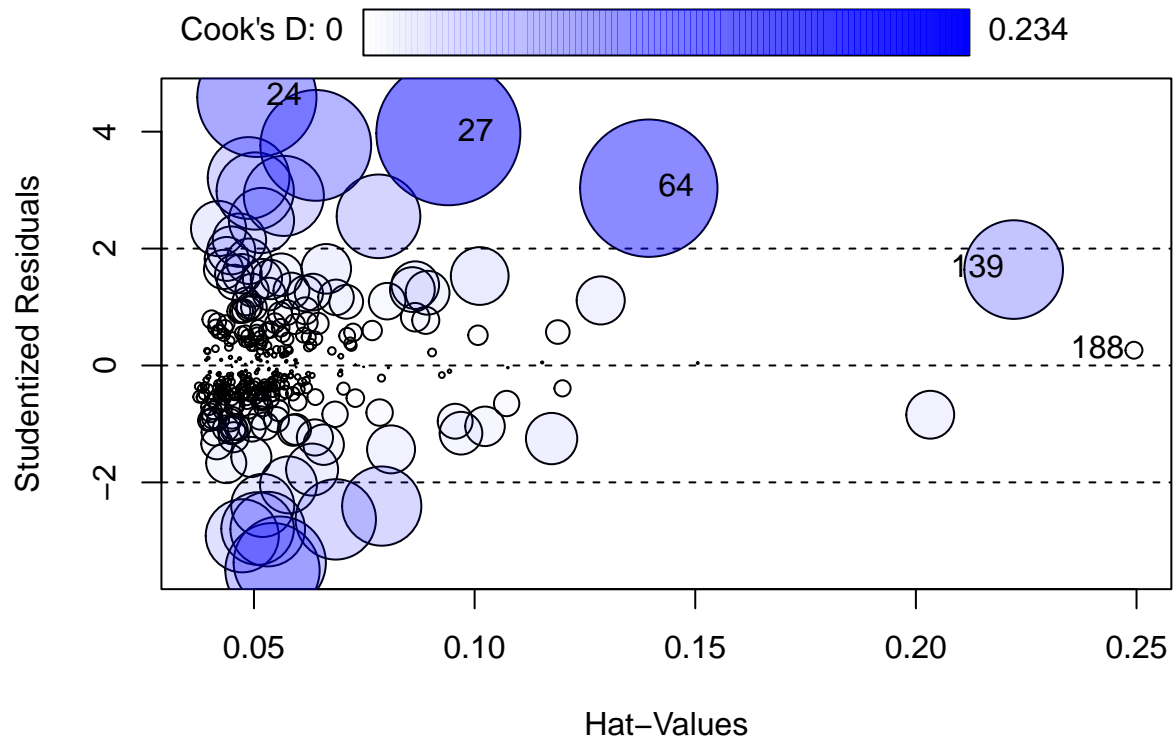
```
##       StudRes        Hat        CookD
## 24   4.5871671 0.05065904 0.160407915
## 27   3.9737485 0.09405150 0.234188155
## 64   3.0331823 0.13947474 0.213025172
## 139  1.6406656 0.22216093 0.109829758
## 188  0.2609713 0.24941692 0.003233066
```

Multicollinearity
```
vif(FinalModel)
```

```
## factor(diversity)    factor(Season)         NO2_NO3              NH4
##          1.026189          1.110467        1.887732         1.171989
##     Nitrification           BiomassC
##          1.123509          1.991204
```

```
performance::check_model(FinalModel, check = c("vif"))
```

Collinearity
High collinearity (VIF) may inflate parameter uncertainty
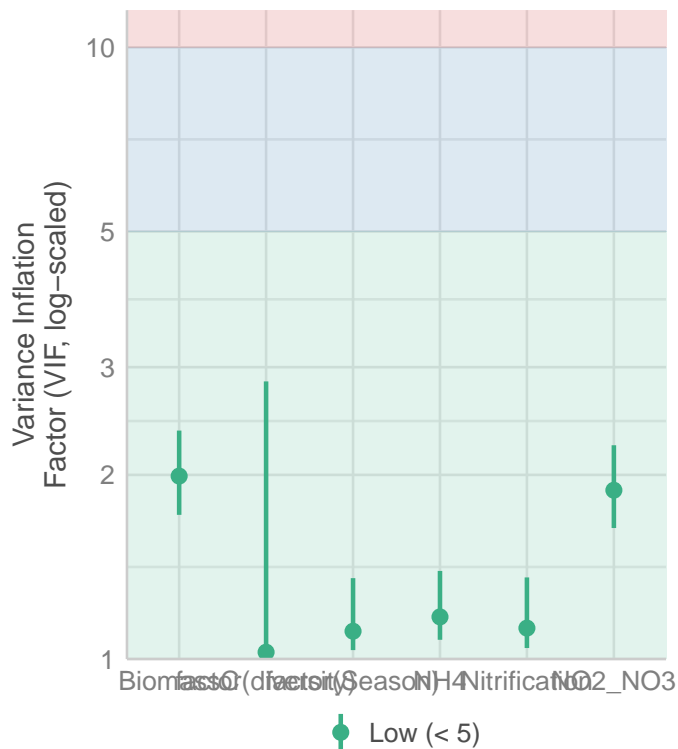

Graphs

```
Ncoef <- fixef(FinalModel)
Ncoef
```

```
##       (Intercept)   factor(diversity)1 factor(Season)Summer
##       -1.9996798            0.1901366           -0.2160541
##           NO2_NO3                  NH4        Nitrification
##         0.3892898            0.2112805            0.1064047
##           BiomassC
##         0.6939279
```

```
success.eq1 <- function(x){Ncoef[1]+Ncoef[2]+
    Ncoef[3]+Ncoef[4]*mean(no.out.tree.l$NO2_NO3)+
    Ncoef[5]*mean(no.out.tree.l$NH4)+Ncoef[6]*mean(no.out.tree.l$Nitrification)+
    Ncoef[7]*x}
success.eq2 <- function(x){Ncoef[1]+
    Ncoef[3]+Ncoef[4]*mean(no.out.tree.l$NO2_NO3)+
    Ncoef[5]*mean(no.out.tree.l$NH4)+Ncoef[6]*mean(no.out.tree.l$Nitrification)+
    Ncoef[7]*x}

ggplot(no.out.tree.l, aes(x=BiomassC, y=BiomassN)) +
  geom_jitter(height = 0.02, width = 0.02, alpha = 0.3, size = 3) +
  stat_function(fun=success.eq1, geom="line", linewidth=0.7, aes(color="High")) +
  stat_function(fun=success.eq2, geom="line", linewidth=0.7, aes(color="Low")) +
  theme_bw() +
```

```
  theme(legend.position = "right") +
  labs(y = "BiomassN", x = "BiomassC") +
  scale_color_viridis("Diversity", discrete = TRUE)
```



```
# Model 3: success
## centering with regard to z-score
tree.c <- tree
tree.c$RockMass_g <- (tree$RockMass_g-mean(tree$RockMass_g))/sd(tree$RockMass_g)
tree.c$RootMass_g <- (tree$RootMass_g-mean(tree$RootMass_g))/sd(tree$RootMass_g)
tree.c$Moisture_g <- (tree$Moisture_g-mean(tree$Moisture_g))/sd(tree$Moisture_g)
tree.c$BulkDensity_g_cm3 <- (tree$BulkDensity_g_cm3-
                             mean(tree$BulkDensity_g_cm3))/sd(tree$BulkDensity_g_cm3)
tree.c$BiomassC <- (tree$BiomassC-mean(tree$BiomassC))/sd(tree$BiomassC)
tree.c$Respiration <- (tree$Respiration-mean(tree$Respiration))/sd(tree$Respiration)
tree.c$NO2_NO3 <- (tree$NO2_NO3-mean(tree$NO2_NO3))/sd(tree$NO2_NO3)
tree.c$NH4 <- (tree$NH4-mean(tree$NH4))/sd(tree$NH4)
tree.c$TIN <- (tree$TIN-mean(tree$TIN))/sd(tree$TIN)
tree.c$BiomassN <- (tree$BiomassN-mean(tree$BiomassN))/sd(tree$BiomassN)
tree.c$Mineralization <- (tree$Mineralization-
                          mean(tree$Mineralization))/sd(tree$Mineralization)
tree.c$Nitrification <- (tree$Nitrification-
                         mean(tree$Nitrification))/sd(tree$Nitrification)
tree.c$DEA <- (tree$DEA-mean(tree$DEA))/sd(tree$DEA)

## Full model
```

```
success.glmer.c <- glmer(success~(1|Site/Plot)+Diversity+BiomassN+BiomassC+RockMass_g+
                         RootMass_g+Moisture_g+Respiration+NH4+Nitrification+DEA+
                         BulkDensity_g_cm3+factor(coresection), family=binomial, data=tree.c)
summary(success.glmer.c)


## Generalized linear mixed model fit by maximum likelihood (Laplace
##    Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: success ~ (1 | Site/Plot) + Diversity + BiomassN + BiomassC +
##     RockMass_g + RootMass_g + Moisture_g + Respiration + NH4 +
##     Nitrification + DEA + BulkDensity_g_cm3 + factor(coresection)
##    Data: tree.c
##
##     AIC      BIC   logLik deviance df.resid
##    44.1    104.7     -6.0     12.1      310
##
## Scaled residuals:
##       Min        1Q    Median        3Q       Max
## -0.084070 -0.015003  0.001657  0.002282  0.005920
##
## Random effects:
##  Groups     Name        Variance  Std.Dev.
##  Plot:Site (Intercept) 1.665e-08 1.29e-04
##  Site      (Intercept) 3.727e+02 1.93e+01
## Number of obs: 326, groups:  Plot:Site, 36; Site, 11
##
## Fixed effects:
##                      Estimate Std. Error z value Pr(>|z|)
## (Intercept)        10.7474696  4.3476548   2.472   0.0134 *
## DiversityLow       -0.0008853  3.2325299   0.000   0.9998
## BiomassN            0.8033754  5.9268509   0.136   0.8922
## BiomassC           -0.0055495  8.0468236  -0.001   0.9994
## RockMass_g         -0.7995148  2.3077320  -0.346   0.7290
## RootMass_g         -0.1894102  2.8381434  -0.067   0.9468
## Moisture_g          1.4796752  5.3434024   0.277   0.7818
## Respiration        -0.0677135  5.0864857  -0.013   0.9894
## NH4                -0.1739761  1.7624236  -0.099   0.9214
## Nitrification       0.0503412  1.8859772   0.027   0.9787
## DEA                -0.4978319  4.5774858  -0.109   0.9134
## BulkDensity_g_cm3   0.1123356  2.3710431   0.047   0.9622
## factor(coresection)2  1.7341823  6.1929224   0.280   0.7795
## factor(coresection)3  2.2576562  6.6721252   0.338   0.7351
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## optimizer (Nelder_Mead) convergence code: 0 (OK)
## Model failed to converge with max|grad| = 0.20008 (tol = 0.002, component 1)


  # error: model failed to converge with max|grad| = 0.203044 (tol = 0.002, component 1)


## Model reduction
success.glmer.c1 <- update(success.glmer.c, .~.-Diversity)
success.glmer.c2 <- update(success.glmer.c1, .~.-Nitrification)
success.glmer.c3 <- update(success.glmer.c2, .~.-BiomassC)
```

```
success.glmer.c4 <- update(success.glmer.c3, .~.-Respiration)
success.glmer.c5 <- update(success.glmer.c4, .~.-BulkDensity_g_cm3)
success.glmer.c6 <- update(success.glmer.c5, .~.-RootMass_g)
success.glmer.c7 <- update(success.glmer.c6, .~.-NH4)
success.glmer.c8 <- update(success.glmer.c7, .~.-DEA) # stop showing error
success.glmer.c9 <- update(success.glmer.c8, .~.-BiomassN) # no significant variables
### Model with variables of interest
success.glmer.c10 <- glmer(success~(1|Site/Plot)+Diversity+BiomassN+BiomassC+
                          RootMass_g+BulkDensity_g_cm3, family=binomial, data=tree.c)
summary(success.glmer.c10)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##    Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: success ~ (1 | Site/Plot) + Diversity + BiomassN + BiomassC +
##     RootMass_g + BulkDensity_g_cm3
##    Data: tree.c
##
##      AIC      BIC   logLik deviance df.resid
##     26.2     56.5     -5.1     10.2      318
##
## Scaled residuals:
##        Min        1Q     Median        3Q       Max
## -0.0177474 -0.0084687  0.0002374  0.0002901  0.0003582
##
## Random effects:
##  Groups     Name        Variance  Std.Dev.
##  Plot:Site (Intercept) 4.908e-07 7.006e-04
##  Site      (Intercept) 8.325e+03 9.124e+01
## Number of obs: 326, groups:  Plot:Site, 36; Site, 11
##
## Fixed effects:
##                    Estimate Std. Error z value Pr(>|z|)
## (Intercept)         16.4607     4.9905   3.298 0.000972 ***
## DiversityLow         0.1215     5.6412   0.022 0.982813
## BiomassN             0.9124     8.9289   0.102 0.918612
## BiomassC            -0.1301     9.2755  -0.014 0.988810
## RootMass_g          -0.2084     3.8935  -0.054 0.957312
## BulkDensity_g_cm3    0.0279     3.2454   0.009 0.993141
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) DvrstL BimssN BimssC RtMss_
## DiversityLw -0.495
## BiomassN     0.062  0.166
## BiomassC     0.171 -0.163 -0.607
## RootMass_g  -0.039 -0.045 -0.168 -0.232
## BlkDnsty__3  0.038 -0.034  0.051  0.043 -0.122
```

```
  # singularity issues without bulk density
  # no significant variables
```

```
## Assumptions
vif(success.glmer.c3)
```

```
##                           GVIF Df GVIF^(1/(2*Df))
## BiomassN              1.837603  1        1.355582
## RockMass_g            1.664039  1        1.289976
## RootMass_g            1.268756  1        1.126391
## Moisture_g            2.182221  1        1.477234
## Respiration           1.914155  1        1.383530
## NH4                   1.116701  1        1.056741
## DEA                   1.648564  1        1.283964
## BulkDensity_g_cm3     1.124398  1        1.060376
## factor(coresection)   2.935006  2        1.308887
```

```
## Coefficients
(success.glmer.c3.logodds <- fixef(success.glmer.c3))
```

```
##          (Intercept)              BiomassN                RockMass_g
##          10.73917268            0.80117394               -0.81418809
##           RootMass_g            Moisture_g               Respiration
##          -0.18956065            1.49072621                0.04478111
##                  NH4                   DEA          BulkDensity_g_cm3
##          -0.16855276           -0.62472908                0.08492009
## factor(coresection)2 factor(coresection)3
##           1.75610788            2.27068975
```

```
(success.glmer.c3.odds <- exp(success.glmer.c3.logodds))
```

```
##          (Intercept)              BiomassN                RockMass_g
##         4.612787e+04          2.228155e+00              4.429989e-01
##           RootMass_g            Moisture_g               Respiration
##         8.273225e-01          4.440319e+00              1.045799e+00
##                  NH4                   DEA          BulkDensity_g_cm3
##         8.448867e-01          5.354065e-01              1.088630e+00
## factor(coresection)2 factor(coresection)3
##         5.789859e+00          9.686079e+00
```

```
(success.glmer.c3.prob <- inv.logit(success.glmer.c3.logodds))
```

```
##          (Intercept)              BiomassN                RockMass_g
##            0.9999783             0.6902255                 0.3069988
##           RootMass_g            Moisture_g               Respiration
##            0.4527512             0.8161872                 0.5111934
##                  NH4                   DEA          BulkDensity_g_cm3
##            0.4579613             0.3487067                 0.5212173
## factor(coresection)2 factor(coresection)3
##            0.8527215             0.9064203
```

```
## graphs
success.eq1 <- function(x){inv.logit(success.glmer.c3.logodds[1]+success.glmer.c3.logodds[2]+
```

```
                                      success.glmer.c3.logodds[3]*x+
                                      success.glmer.c3.logodds[4]*mean(tree.c$BiomassC)+
                                      success.glmer.c3.logodds[5]*mean(tree.c$RootMass_g)+
                                      success.glmer.c3.logodds[6]*mean(tree.c$BulkDensity_g_cm3))}

gg1 <- ggplot(tree.c, aes(x=BiomassN, y=success)) +
  geom_jitter(height = 0.02, width = 0.02, alpha = 0.3, size = 3) +
  stat_function(fun=success.eq1, geom="line", linewidth=0.7, color="blue") +
  theme_bw() +
  labs(y = "Probability of afforestation success",
       x = "Biomass N, centered by z-score")

## graph 2 (glm)
gg2 <- ggplot(tree.c, aes(x=BiomassN, y=success)) +
  geom_jitter(height = 0.02, width = 0.02, alpha = 0.3, size = 3) +
  geom_line(stat = "smooth", method = "glm", method.args = list(family = "binomial"),
            size = 1.1, color="purple") +
  geom_ribbon(stat="smooth", method = "glm", se = TRUE, alpha = 0.1,
              method.args = list(family = "binomial")) +
  ylab("Probability of afforestation success") +
  xlab("Biomass N, centered by z-score") +
  theme_bw()
```

Appendix 2 Code

```r
knitr::opts_chunk$set(echo = TRUE, eval = TRUE, warning = FALSE, message = FALSE)
# Load dataset and packages
tree <- read.csv("C:\\Users\\Siyu Sun\\Documents\\ENV710_Project\\Data\\MTNYCData_modified2_noDEAna.csv
# tree <- read.csv("MTNYCData_modified2_noDEAna.csv")
pacman::p_load(dplyr, ggplot2, lme4, lmerTest, lmtest, GGally, car, DHARMa, boot,
               viridis, ggpubr, performance, MuMIn, merTools, lattice)


# Data exploration
ggpairs(tree[,c(9:17)])
ggpairs(tree[,c(18:25)])


# Model 1: Biomass C
## Full model
lm1 <- lmer(log(BiomassC) ~ factor(Success) + factor(Diversity) + factor(Season) +
               log(Respiration) + log(BiomassN + 6.53271102 + 0.001) +
               log(NO2_NO3 + 5.9671545 + 0.001) + log(NH4 + 2.422020695 + 0.001) +
               log(RootMass_g + 0.782 + 0.001) + log(Moisture_g) + factor(coresection) +
               factor(Year) + (1|Site/Plot), data = tree)
summary(lm1)


## Model reduction
lm2 <- update(lm1, ~.-log(RootMass_g + 0.782 + 0.001))
# summary(lm2)
lm3 <- update(lm2, ~.-factor(Year))
# summary(lm3)
lm4 <- update(lm3, ~.-log(NO2_NO3 + 5.9671545 + 0.001))
# summary(lm4)
lm5 <- update(lm4, ~.-factor(Success))
# summary(lm5)
lm6 <- update(lm5, ~.-log(NH4 + 2.422020695 + 0.001))
summary(lm6)


### delete the outlier according to the outlier plot
tree1 <- tree[-c(33),]
lm7 <- lmer(log(BiomassC) ~ factor(Diversity) + factor(Season) + log(Respiration) +
               log(BiomassN + 4.79350047 + 0.001) + log(Moisture_g) + factor(coresection) +
               (1 | Site/Plot), data = tree1)
summary(lm7)


AIC(lm1, lm2, lm3,lm4,lm5,lm6, lm7)



## Assumptions
### dept vars normality
shapiro.test(tree$BiomassC)
shapiro.test(log(tree$BiomassC))
### Residuals mean & normality
mean(residuals(lm6))
shapiro.test(residuals(lm6))
lmm6 <- simulateResiduals(fittedModel = lm6)
plot(lmm6)
```

```r
check_model(lm6, check = c("reqq", "qq", "linearity"))
### Goodness of fit / R2 FE RE
r.squaredGLMM(lm6)
plotFEsim(FEsim(lm6), intercept = T)
dotplot(ranef(lm6))
### Outlier test
outlierTest(lm6)
influencePlot(lm6)
### Multicollinearity
vif(lm6)
check_model(lm6, check = c("vif"))


## Graphs
### Respiration-diversity
a <- fixef(lm7)

eq.re <- function(x){exp(a[1]+a[2]+a[4]*log(x)+ a[5]*mean(log(tree1$BiomassN +
                                                    4.79350047 + 0.001))+
                        a[6]*mean(log(tree1$Moisture_g)))}

eq.re1 <- function(x){exp(a[1]+a[4]*log(x)+ a[5]*mean(log(tree1$BiomassN +
                                                    4.79350047 + 0.001))+
                        a[6]*mean(log(tree1$Moisture_g)))}

library(viridis)
ggplot(tree1, aes(x=Respiration, y=BiomassC)) +
  geom_jitter(height = 0.02, width = 0.02, alpha = 0.3, size = 3) +
  stat_function(fun=eq.re, geom="line", linewidth=0.7, aes(color="Low")) +
  stat_function(fun=eq.re1, geom="line", linewidth=0.7, aes(color="High")) +
  theme_bw() +
  theme(legend.position = "right") +
  labs(y = "BiomassC", x = "Respiration") +
  scale_color_viridis("Diversity", discrete = TRUE)

## Moisture-diversity
a <- fixef(lm7)

eq.mo <- function(x){exp(a[1]+a[2]+a[4]*mean(log(tree1$Respiration))+
                        a[5]*mean(log(tree1$BiomassN + 4.79350047 + 0.001))+
                        a[6]*log(x))}

eq.mo1 <- function(x){exp(a[1]+a[4]*mean(log(tree1$Respiration))+
                        a[5]*mean(log(tree1$BiomassN + 4.79350047 + 0.001))+
                        a[6]*log(x))}

ggplot(tree1, aes(x=Moisture_g, y=BiomassC)) +
  geom_jitter(height = 0.02, width = 0.02, alpha = 0.3, size = 3) +
  stat_function(fun=eq.mo, geom="line", linewidth=0.7, aes(color="Low")) +
  stat_function(fun=eq.mo1, geom="line", linewidth=0.7, aes(color="High")) +
  theme_bw() +
  theme(legend.position = "right") +
  labs(y = "BiomassC", x = "Moisture") +
  scale_color_viridis("Diversity", discrete = TRUE)
```

```r
# Calculate mean and standard error for each year
mean_data <- aggregate(tree1$BiomassC, by=list(group=tree1$coresection), FUN=mean)
names(mean_data)[2] <- "Mean"
se_data <- aggregate(tree1$BiomassC, by=list(group=tree1$coresection), FUN=sd)
se_data$se <- se_data$x / sqrt(length(tree1$BiomassC))

# Create bar plot with error bars
ggplot(mean_data, aes(x=factor(group), y=Mean, fill=factor(group))) +
  geom_bar(stat="identity", position="dodge", alpha = 0.5) +
  geom_errorbar(aes(ymin=Mean-se_data$se, ymax=Mean+se_data$se),width=.2) +
  labs(x="Core Section Groups", y="Biomass",fill = "Core Section") +
  scale_fill_discrete(labels=c("0-10cm","10-30cm","above 30cm")) +
  theme_bw()
# Model 2: Biomass N
## log-transform all of the variables
tree.l <- tree
for (i in 1:326){
  tree.l$RockMass_g[i] <- log(tree$RockMass_g[i] + abs(min(tree$RockMass_g)) + 0.001)
  tree.l$RootMass_g[i] <- log(tree$RootMass_g[i] + abs(min(tree$RootMass_g)) + 0.001)
  tree.l$Moisture_g[i] <- log(tree$Moisture_g[i] + abs(min(tree$Moisture_g)) + 0.001)
  tree.l$BulkDensity_g_cm3[i] <- log(tree$BulkDensity_g_cm3[i] +
                                      abs(min(tree$BulkDensity_g_cm3)) + 0.001)
  tree.l$BiomassC[i] <- log(tree$BiomassC[i] + abs(min(tree$BiomassC)) + 0.001)
  tree.l$Respiration[i] <- log(tree$Respiration[i] + abs(min(tree$Respiration)) + 0.001)
  tree.l$NO2_NO3[i] <- log(tree$NO2_NO3[i] + abs(min(tree$NO2_NO3)) + 0.001)
  tree.l$NH4[i] <- log(tree$NH4[i] + abs(min(tree$NH4)) + 0.001)
  tree.l$TIN[i] <- log(tree$TIN[i] + abs(min(tree$TIN)) + 0.001)
  tree.l$BiomassN[i] <- log(tree$BiomassN[i] + abs(min(tree$BiomassN)) + 0.001)
  tree.l$DEA[i] <- log(tree$DEA[i] + abs(min(tree$DEA)) + 0.001)
}


### Remove highly influencial observations
no.out.tree.l <- tree.l[-c(33,29,193),]


## Full model
noY.finalm1 <- lmer(BiomassN ~ Success + factor(diversity) + factor(Season) +
                    Respiration + NO2_NO3 + NH4 + Nitrification + RootMass_g +
                    Moisture_g + factor(coresection) + BiomassC +
                    (1|Site/Plot), data = no.out.tree.l)
summary(noY.finalm1)


## Model reduction
noY.finalm2 <- update(noY.finalm1,.~.-Respiration)
summary(noY.finalm2)


noY.finalm3 <- update(noY.finalm2,.~.-Success)
summary(noY.finalm3)


noY.finalm4 <- update(noY.finalm3,.~.-factor(coresection))
summary(noY.finalm4)


noY.finalm5 <- update(noY.finalm4,.~.-Moisture_g)
summary(noY.finalm5)
```

```r
FinalModel <- update(noY.finalm5,.~.-RootMass_g)

AIC(noY.finalm1, noY.finalm2, noY.finalm3, noY.finalm4, noY.finalm5, FinalModel)

## Assumptions
### Residuals mean & normality
mean(residuals(FinalModel))
shapiro.test(residuals(FinalModel))
library(DHARMa)
lmm6 <- simulateResiduals(fittedModel = FinalModel)
plot(lmm6)
library(performance)
performance::check_model(FinalModel, check = c("reqq", "qq", "linearity"))
### Goodness of fit / R2 FE RE
library(merTools)
merTools::plotFEsim(FEsim(FinalModel), intercept = T)
library(lattice)
dotplot(ranef(FinalModel))
### Outlier test
library(car)
outlierTest(FinalModel)
influencePlot(FinalModel)
### Multicollinearity
vif(FinalModel)
performance::check_model(FinalModel, check = c("vif"))

## Graphs
Ncoef <- fixef(FinalModel)
Ncoef

success.eq1 <- function(x){Ncoef[1]+Ncoef[2]+
    Ncoef[3]+Ncoef[4]*mean(no.out.tree.l$NO2_NO3)+
    Ncoef[5]*mean(no.out.tree.l$NH4)+Ncoef[6]*mean(no.out.tree.l$Nitrification)+
    Ncoef[7]*x}
success.eq2 <- function(x){Ncoef[1]+
    Ncoef[3]+Ncoef[4]*mean(no.out.tree.l$NO2_NO3)+
    Ncoef[5]*mean(no.out.tree.l$NH4)+Ncoef[6]*mean(no.out.tree.l$Nitrification)+
    Ncoef[7]*x}

ggplot(no.out.tree.l, aes(x=BiomassC, y=BiomassN)) +
  geom_jitter(height = 0.02, width = 0.02, alpha = 0.3, size = 3) +
  stat_function(fun=success.eq1, geom="line", linewidth=0.7, aes(color="High")) +
  stat_function(fun=success.eq2, geom="line", linewidth=0.7, aes(color="Low")) +
  theme_bw() +
  theme(legend.position = "right") +
  labs(y = "BiomassN", x = "BiomassC") +
  scale_color_viridis("Diversity", discrete = TRUE)

# Model 3: success
## centering with regard to z-score
tree.c <- tree
tree.c$RockMass_g <- (tree$RockMass_g-mean(tree$RockMass_g))/sd(tree$RockMass_g)
tree.c$RootMass_g <- (tree$RootMass_g-mean(tree$RootMass_g))/sd(tree$RootMass_g)
```

```r
tree.c$Moisture_g <- (tree$Moisture_g-mean(tree$Moisture_g))/sd(tree$Moisture_g)
tree.c$BulkDensity_g_cm3 <- (tree$BulkDensity_g_cm3-
                            mean(tree$BulkDensity_g_cm3))/sd(tree$BulkDensity_g_cm3)
tree.c$BiomassC <- (tree$BiomassC-mean(tree$BiomassC))/sd(tree$BiomassC)
tree.c$Respiration <- (tree$Respiration-mean(tree$Respiration))/sd(tree$Respiration)
tree.c$NO2_NO3 <- (tree$NO2_NO3-mean(tree$NO2_NO3))/sd(tree$NO2_NO3)
tree.c$NH4 <- (tree$NH4-mean(tree$NH4))/sd(tree$NH4)
tree.c$TIN <- (tree$TIN-mean(tree$TIN))/sd(tree$TIN)
tree.c$BiomassN <- (tree$BiomassN-mean(tree$BiomassN))/sd(tree$BiomassN)
tree.c$Mineralization <- (tree$Mineralization-
                          mean(tree$Mineralization))/sd(tree$Mineralization)
tree.c$Nitrification <- (tree$Nitrification-
                         mean(tree$Nitrification))/sd(tree$Nitrification)
tree.c$DEA <- (tree$DEA-mean(tree$DEA))/sd(tree$DEA)

## Full model
success.glmer.c <- glmer(success~(1|Site/Plot)+Diversity+BiomassN+BiomassC+RockMass_g+
                         RootMass_g+Moisture_g+Respiration+NH4+Nitrification+DEA+
                         BulkDensity_g_cm3+factor(coresection), family=binomial, data=tree.c)
summary(success.glmer.c)
  # error: model failed to converge with max|grad| = 0.203044 (tol = 0.002, component 1)

## Model reduction
success.glmer.c1 <- update(success.glmer.c, .~.-Diversity)
success.glmer.c2 <- update(success.glmer.c1, .~.-Nitrification)
success.glmer.c3 <- update(success.glmer.c2, .~.-BiomassC)
success.glmer.c4 <- update(success.glmer.c3, .~.-Respiration)
success.glmer.c5 <- update(success.glmer.c4, .~.-BulkDensity_g_cm3)
success.glmer.c6 <- update(success.glmer.c5, .~.-RootMass_g)
success.glmer.c7 <- update(success.glmer.c6, .~.-NH4)
success.glmer.c8 <- update(success.glmer.c7, .~.-DEA) # stop showing error
success.glmer.c9 <- update(success.glmer.c8, .~.-BiomassN) # no significant variables
### Model with variables of interest
success.glmer.c10 <- glmer(success~(1|Site/Plot)+Diversity+BiomassN+BiomassC+
                           RootMass_g+BulkDensity_g_cm3, family=binomial, data=tree.c)
summary(success.glmer.c10)
  # singularity issues without bulk density
  # no significant variables

## Assumptions
vif(success.glmer.c3)

## Coefficients
(success.glmer.c3.logodds <- fixef(success.glmer.c3))
(success.glmer.c3.odds <- exp(success.glmer.c3.logodds))
(success.glmer.c3.prob <- inv.logit(success.glmer.c3.logodds))

## graphs
success.eq1 <- function(x){inv.logit(success.glmer.c3.logodds[1]+success.glmer.c3.logodds[2]+
                                      success.glmer.c3.logodds[3]*x+
                                      success.glmer.c3.logodds[4]*mean(tree.c$BiomassC)+
                                      success.glmer.c3.logodds[5]*mean(tree.c$RootMass_g)+
                                      success.glmer.c3.logodds[6]*mean(tree.c$BulkDensity_g_cm3))}
```

```
gg1 <- ggplot(tree.c, aes(x=BiomassN, y=success)) +
  geom_jitter(height = 0.02, width = 0.02, alpha = 0.3, size = 3) +
  stat_function(fun=success.eq1, geom="line", linewidth=0.7, color="blue") +
  theme_bw() +
  labs(y = "Probability of afforestation success",
       x = "Biomass N, centered by z-score")

## graph 2 (glm)
gg2 <- ggplot(tree.c, aes(x=BiomassN, y=success)) +
  geom_jitter(height = 0.02, width = 0.02, alpha = 0.3, size = 3) +
  geom_line(stat = "smooth", method = "glm", method.args = list(family = "binomial"),
            size = 1.1, color="purple") +
  geom_ribbon(stat="smooth", method = "glm", se = TRUE, alpha = 0.1,
              method.args = list(family = "binomial")) +
  ylab("Probability of afforestation success") +
  xlab("Biomass N, centered by z-score") +
  theme_bw()
```