Basal area

Tim White

June 2024

Packages and data

```r
library(car)
library(tidyverse)

data_stems <- read_csv("data/data_stems.csv") %>%
              mutate(plot_id = as.factor(plot_id),
                     size_class = as.factor(size_class),
                     harvested = as.factor(harvested),
                     vegetation_type = as.factor(vegetation_type),
                     milpa = as.factor(milpa))
data_plots <- read_csv("data/data_plots.csv") %>%
              mutate(plot_id = as.factor(plot_id),
                     harvested = as.factor(harvested),
                     vegetation_type = as.factor(vegetation_type),
                     milpa = as.factor(milpa))

plot_theme <- theme(plot.background = element_rect(fill = "white"),
                    plot.title = element_blank(),
                    plot.subtitle = element_text(family="sans", face="plain"),
                    axis.title.x = element_text(family="sans", face="bold"),
                    axis.title.y = element_text(family="sans", face="bold"),
                    axis.text.x = element_text(family="sans", face="plain"),
                    axis.text.y = element_text(family="sans", face="plain"),
                    panel.background = element_rect(fill="white"),
                    panel.grid.major.x = element_blank(),
                    panel.grid.major.y = element_line(color="gainsboro"),
                    panel.grid.minor = element_blank(),
                    axis.ticks = element_blank(),
                    legend.background = element_rect(color="black", fill = "white"),
                    legend.position = c(0.9, 0.85),
                    legend.title = element_text(color = "black", face = "bold", hjust = 0.5),
                    legend.text = element_text(color = "black"))
```
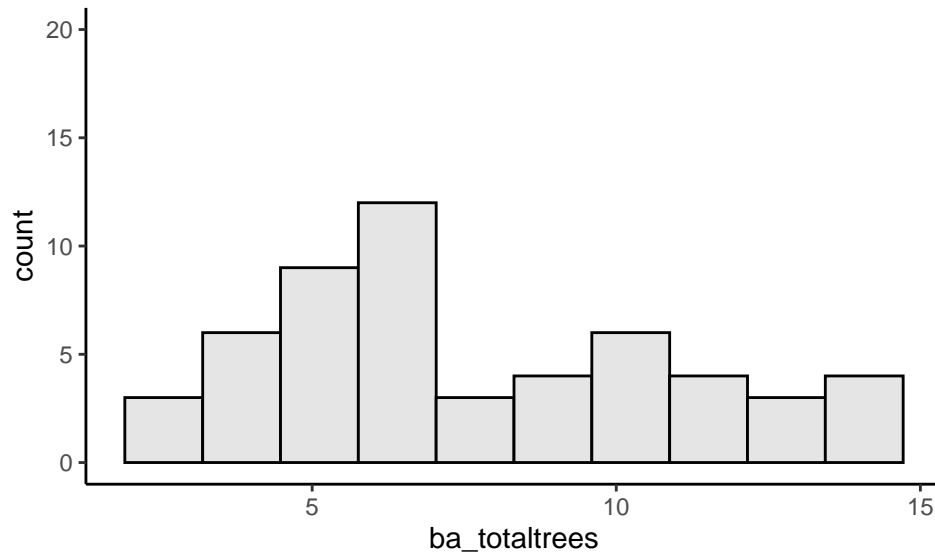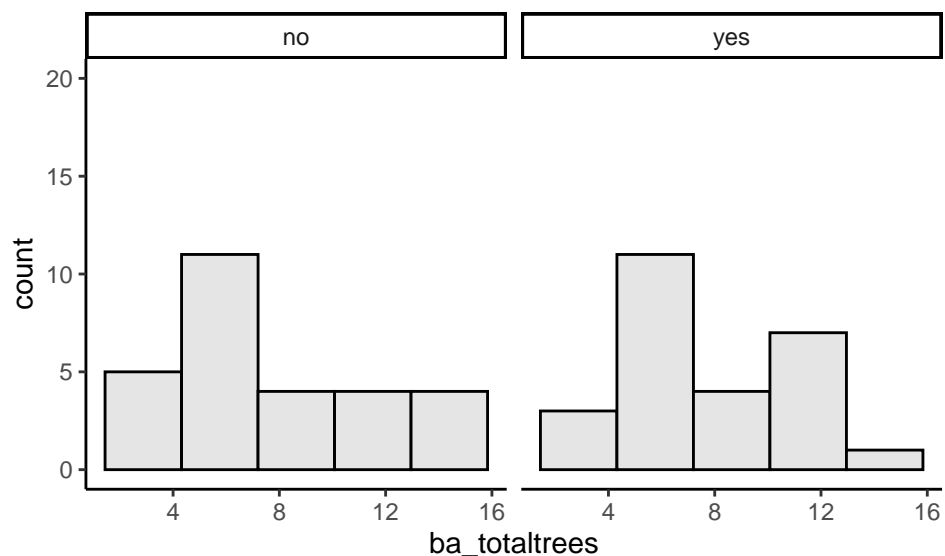
Figures

Histograms of total basal area of trees

```r
# All plots
data_plots %>%
  ggplot() +
    geom_histogram(aes(x = ba_totaltrees), bins = 10, fill = "gray90", col = "black") +
    theme_classic() + ylim(c(0,20))
```



```r
# Harvested vs. unharvested plots
data_plots %>%
  ggplot() +
  geom_histogram(aes(x = ba_totaltrees), bins = 5, fill = "gray90", col = "black") +
  facet_wrap(~harvested) +
  theme_classic() + ylim(c(0,20))
```

Five-number summaries of total basal area of trees

```
# All plots
summary(data_plots$ba_totaltrees)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   2.498   5.161   6.639   7.660  10.316  14.015
```

```
# Harvested vs. unharvested plots
data_plots %>% filter(harvested == "yes") %>% pull(ba_totaltrees) %>% summary()
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   2.788   4.860   6.506   7.550  10.248  13.785
```

```
data_plots %>% filter(harvested == "no") %>% pull(ba_totaltrees) %>% summary()
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   2.498   5.450   6.765   7.763  10.568  14.015
```

Box plots of total basal area of trees

```
# By harvesting status
ba_fig1 <- data_plots %>%
  mutate(harvested = fct_recode(harvested,
                                "Harvested" = "yes", "Unharvested" = "no")) %>%
  ggplot() +
    geom_boxplot(aes(x = fct_relevel(harvested, "Harvested", "Unharvested"),
                     y = ba_totaltrees),
                 fill = "mistyrose2", outlier.shape = 21,
                 outlier.color = "gray20", outlier.fill = "gray80") +
    plot_theme +
    coord_cartesian(ylim = c(0, 15)) +
    labs(x = "Plot status", y = "Total basal area of trees (m2 per hectare)")
ggsave("figures/ba_fig1.png", height = 4, width = 6)
ba_fig1
```
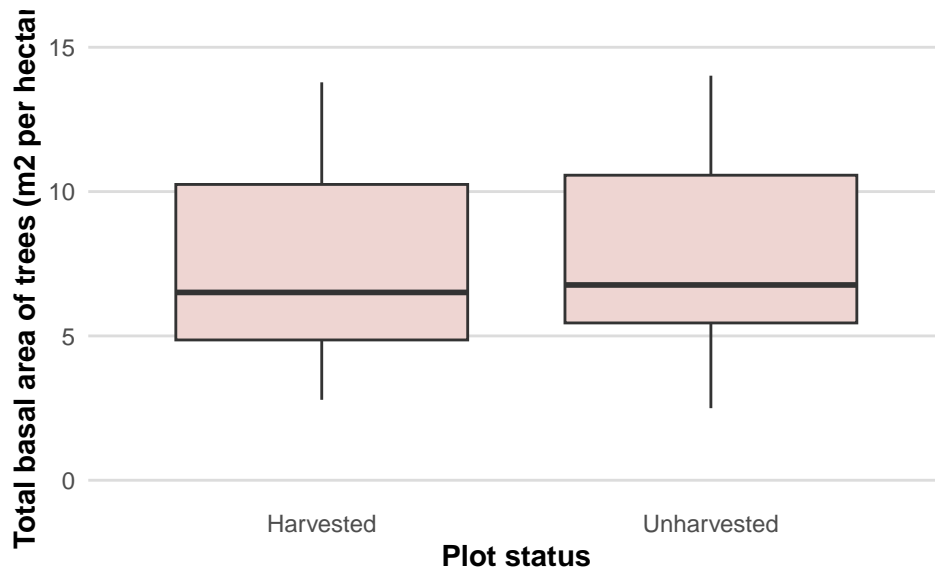
```r
# By vegetation type
ba_fig2 <- data_plots %>%
  mutate(vegetation_type = fct_recode(vegetation_type,
                                      "Ju'uche'" = "juuche",
                                      "Keelenche'" = "keelenche",
                                      "Nuku'uch che'" = "nukuuchche")) %>%
  ggplot() +
  geom_boxplot(aes(x = vegetation_type, y = ba_totaltrees),
               fill = "mistyrose2", outlier.shape = 21,
               outlier.color = "gray20", outlier.fill = "gray80") +
  plot_theme +
  coord_cartesian(ylim = c(0, 15)) +
  labs(x = "Vegetation type", y = "Total basal area of trees (m2 per hectare)")
ggsave("figures/ba_fig2.png", height = 4, width = 6)
ba_fig2
```
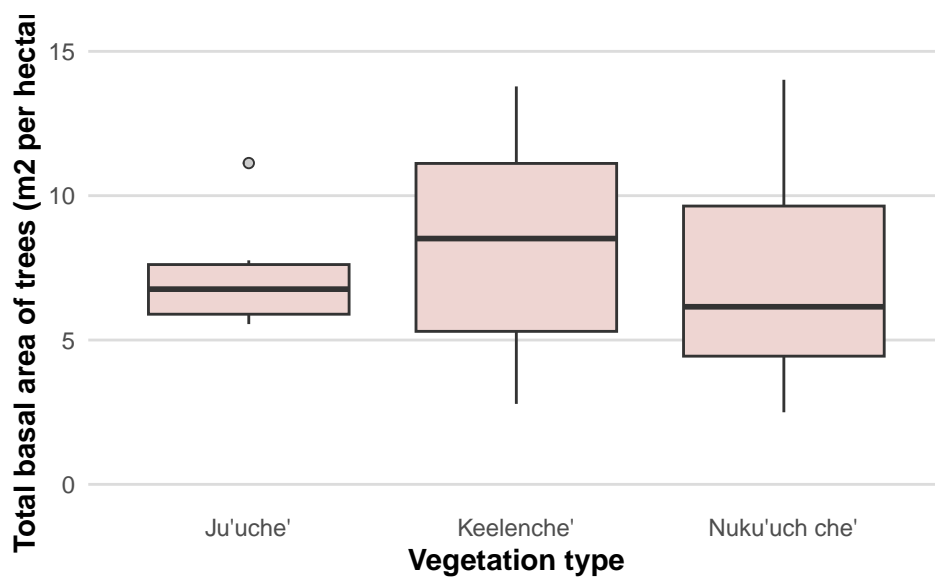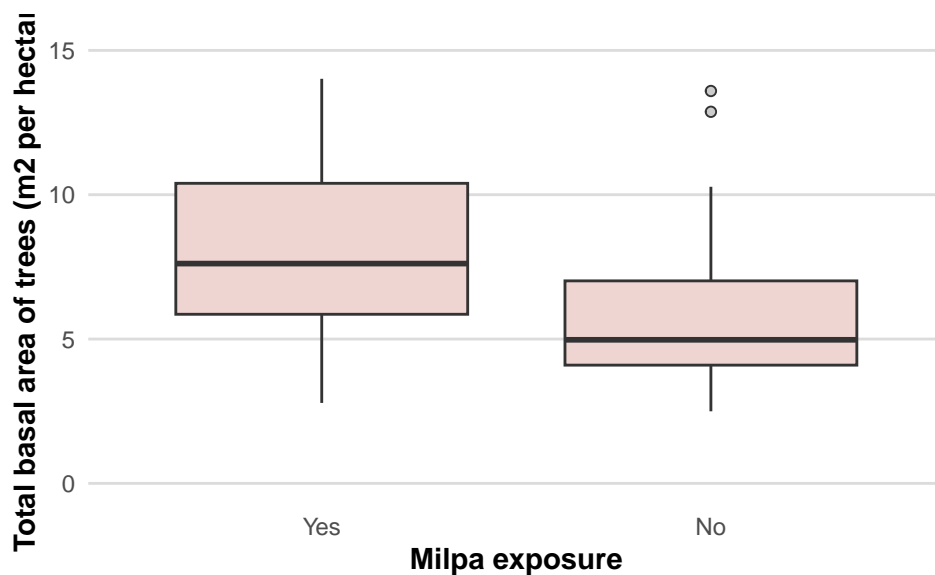
```
# By milpa
ba_fig3 <- data_plots %>%
  mutate(milpa = fct_recode(milpa,
                            "Yes" = "yes", "No" = "no")) %>%
  ggplot() +
  geom_boxplot(aes(x = fct_relevel(milpa, "Yes", "No"), y = ba_totaltrees),
               fill = "mistyrose2", outlier.shape = 21,
               outlier.color = "gray20", outlier.fill = "gray80") +
  plot_theme +
  coord_cartesian(ylim = c(0, 15)) +
  labs(x = "Milpa exposure", y = "Total basal area of trees (m2 per hectare)")
ggsave("figures/ba_fig3.png", height = 4, width = 6)
ba_fig3
```
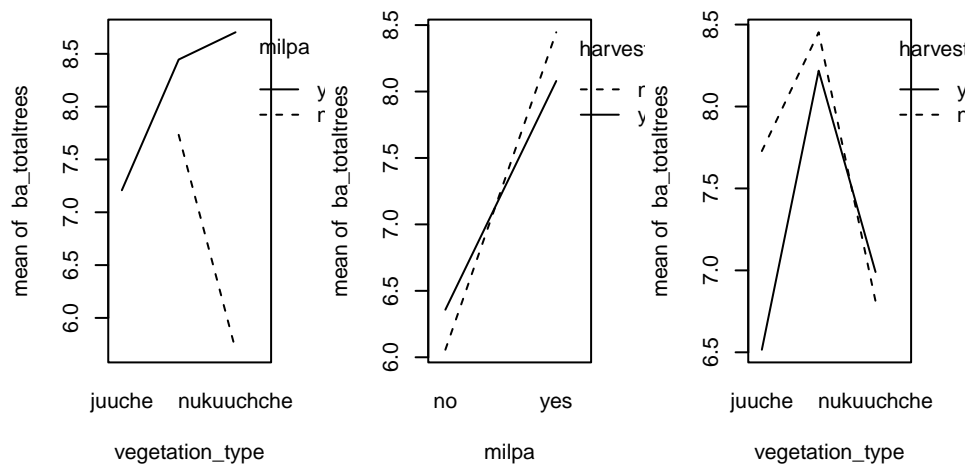

Interaction plots

```
par(mfrow = c(1,3))
with(data_plots, {interaction.plot(vegetation_type, milpa, ba_totaltrees)
                  interaction.plot(milpa, harvested, ba_totaltrees)
                  interaction.plot(vegetation_type, harvested, ba_totaltrees)})
```
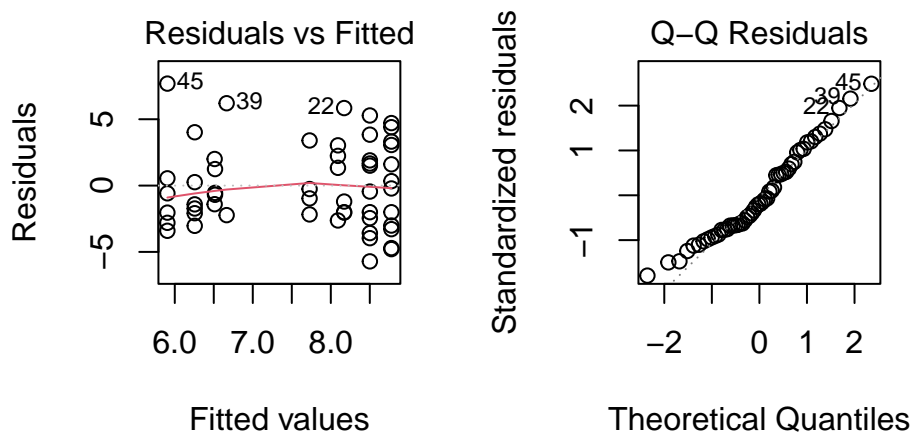
Models

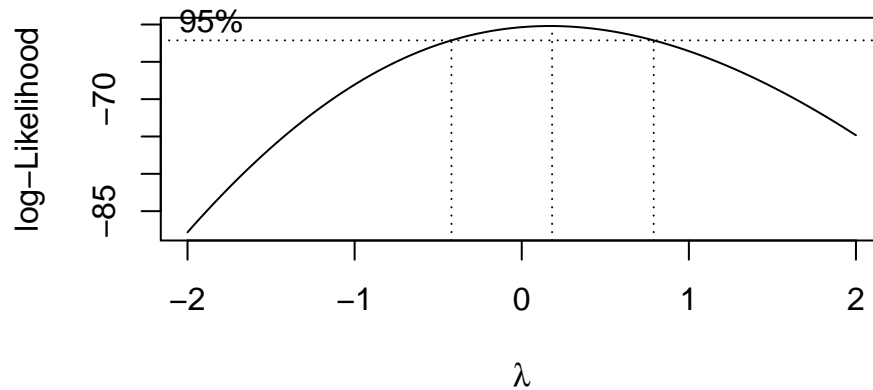### Fit model and check assumptions

```r
mod <- lm(ba_totaltrees ~ harvested + vegetation_type + milpa +
            harvested:vegetation_type + harvested:milpa, data = data_plots)

par(mfrow = c(1,2))
plot(mod, which = 1:2)
```
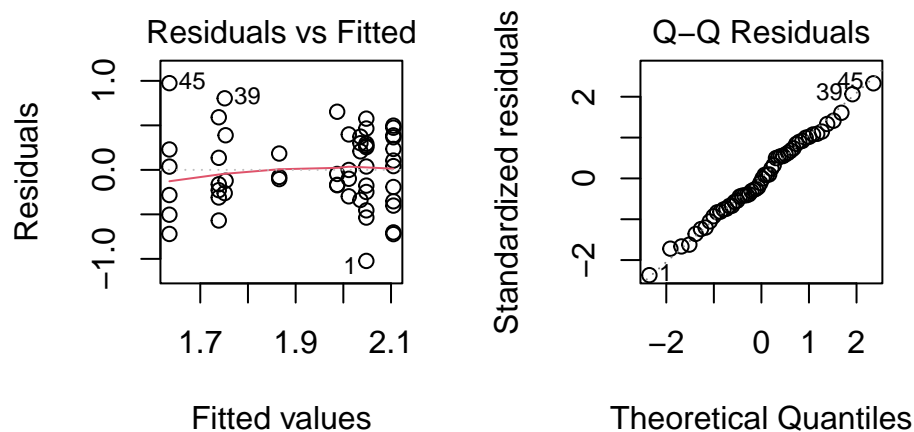

Attempt transformation of response

```r
# Residuals vs. fitted looks pretty good, but we can try Box-Cox just to be safe
MASS::boxcox(mod)
```

```r
# Box-Cox suggests log transformation; re-fit model with log response
mod.transform <- lm(log(ba_totaltrees) ~ harvested + vegetation_type + milpa +
                      harvested:vegetation_type + harvested:milpa, data = data_plots)

# Check constant variance and normality assumptions
par(mfrow = c(1,2))
plot(mod.transform, which = 1:2)
```


Assumptions look about the same. For the sake of interpretability, we decide to go with the untransformed model.

Summarize selected model and run ANOVA

```r
# Check for outliers
outlierTest(mod)
```

```
## No Studentized residuals with Bonferroni p < 0.05
## Largest |rstudent|:
##    rstudent unadjusted p-value Bonferroni p
## 45 2.644258         0.011231      0.60648
```

```
# No significant outliers; look at model summary
summary(mod)
```

```
##
## Call:
## lm(formula = ba_totaltrees ~ harvested + vegetation_type + milpa +
##     harvested:vegetation_type + harvested:milpa, data = data_plots)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.7134 -2.1526 -0.5669  1.9777  7.6870
##
## Coefficients:
##                                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)                            5.4625     2.3386   2.336   0.0239 *
## harvestedyes                          -0.7826     3.4463  -0.227   0.8214
## vegetation_typekeelenche               1.0486     1.9027   0.551   0.5842
## vegetation_typenukuuchche              0.4430     2.2029   0.201   0.8415
## milpayes                               2.2657     1.6420   1.380   0.1743
## harvestedyes:vegetation_typekeelenche  0.9365     2.8697   0.326   0.7456
## harvestedyes:vegetation_typenukuuchche 1.1330     3.2612   0.347   0.7299
## harvestedyes:milpayes                 -0.4297     2.3252  -0.185   0.8542
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.331 on 46 degrees of freedom
## Multiple R-squared:  0.1105, Adjusted R-squared:  -0.02484
## F-statistic: 0.8165 on 7 and 46 DF,  p-value: 0.5784
```

```
# Construct ANOVA table (use type 2 SS since data are unbalanced)
Anova(mod, type = 2)
```

```
## Anova Table (Type II tests)
##
## Response: ba_totaltrees
##                           Sum Sq Df F value  Pr(>F)
## harvested                   0.51  1  0.0458 0.83146
## vegetation_type            12.31  2  0.5550 0.57788
## milpa                      34.54  1  3.1134 0.08429 .
## harvested:vegetation_type   1.45  2  0.0655 0.93664
## harvested:milpa             0.38  1  0.0341 0.85421
## Residuals                 510.25 46
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We find that no predictors are significant.

Pairwise comparisons

```r
# Confirm lack of significance of harvested using Tukey HSD
# Unharvested plots have higher BA than harvested plots, but difference is not significant
TukeyHSD(aov(mod), "harvested", conf.level = 0.95)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = mod)
##
## $harvested
##              diff       lwr      upr     p adj
## yes-no -0.2136882 -2.039544 1.612168 0.8148056
```