

Stem density

Tim White

June 2024

Packages and data

```
library(car)
library(tidyverse)

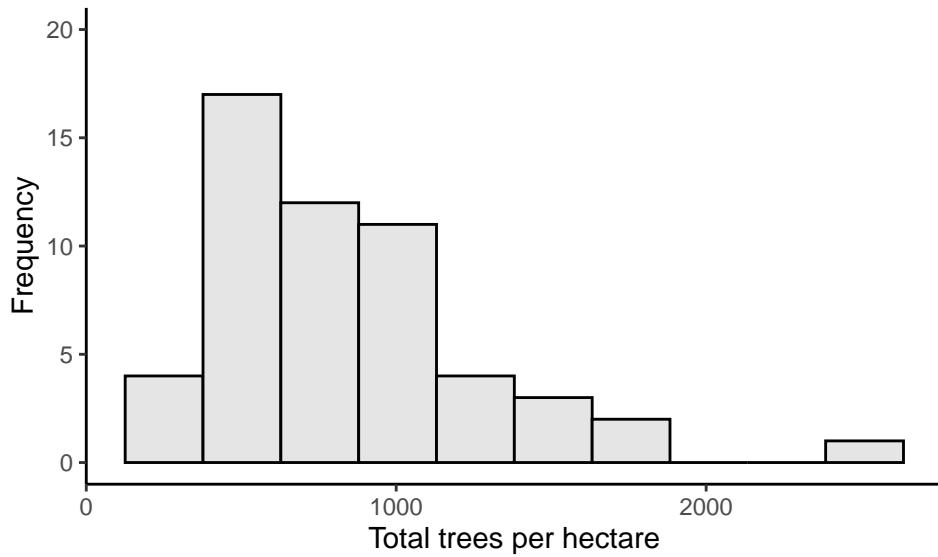
data_stems <- read_csv("data/data_stems.csv") %>%
  mutate(plot_id = as.factor(plot_id),
        size_class = as.factor(size_class),
        harvested = as.factor(harvested),
        vegetation_type = as.factor(vegetation_type),
        milpa = as.factor(milpa))
data_plots <- read_csv("data/data_plots.csv")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

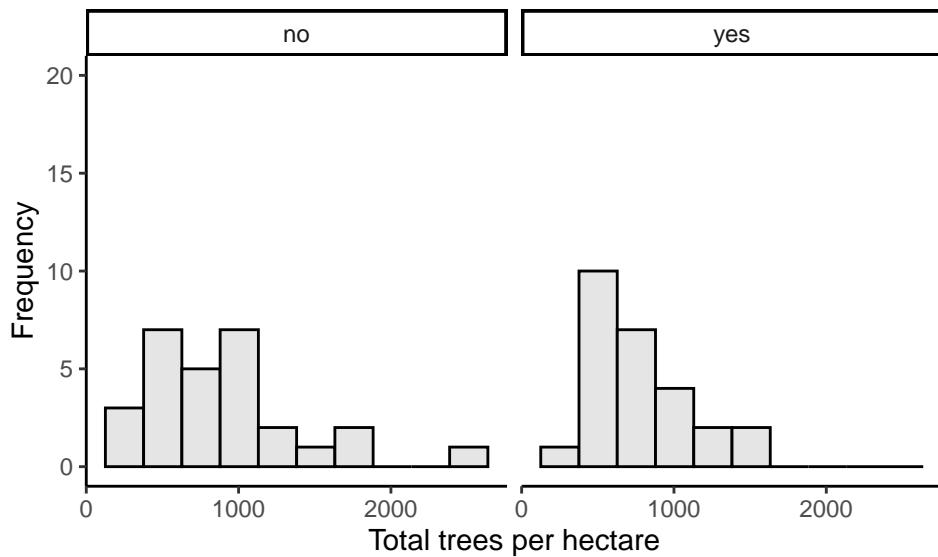
Figures

Histograms of total trees per hectare

```
# All plots
data_plots %>%
  ggplot() +
  geom_histogram(aes(x = stemden_totaltrees), bins = 10, fill = "gray90", col = "black") +
  theme_classic() + ylim(c(0,20)) + labs(x = "Total trees per hectare", y = "Frequency")
```



```
# Harvested vs. unharvested plots
data_plots %>%
  ggplot() +
  geom_histogram(aes(x = stemden_totaltrees), bins = 10, fill = "gray90", col = "black") +
  facet_wrap(~harvested) +
  theme_classic() + ylim(c(0,20)) + labs(x = "Total trees per hectare", y = "Frequency")
```



Five-number summaries of total trees per hectare

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

##      Min. 1st Qu. Median     Mean 3rd Qu.     Max.
##    254.6   557.0  779.9  848.8 1018.6 2514.6

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

##      Min. 1st Qu. Median     Mean 3rd Qu.     Max.
##    254.6   517.3  716.2  778.6  923.1 1496.1

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

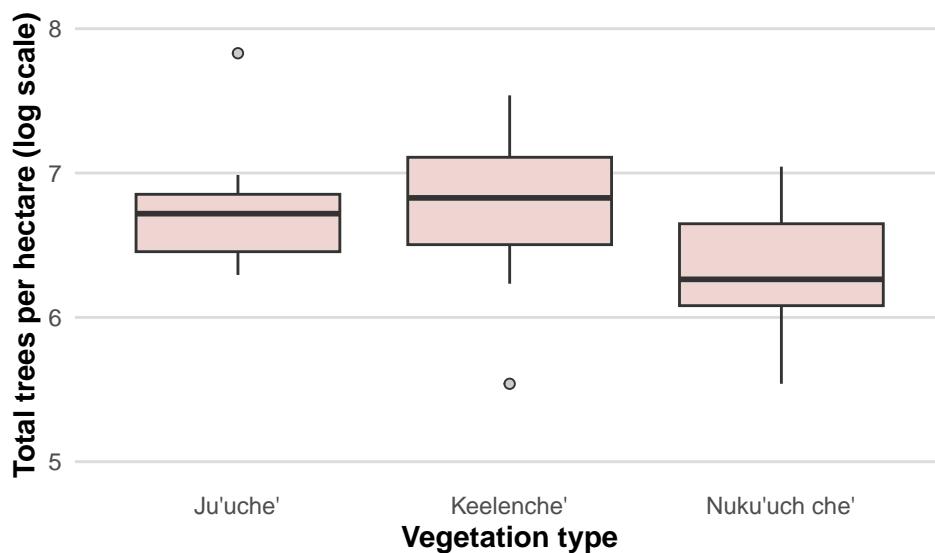
##      Min. 1st Qu. Median     Mean 3rd Qu.     Max.
##    254.6   604.8  827.6  914.0 1034.5 2514.6
```

Box plots of total trees per hectare (log scale)

```
# By harvesting status
stemden_fig1 <- data_plots %>%
  mutate(harvested = fct_recode(harvested,
                                "Harvested" = "yes", "Unharvested" = "no")) %>%
  ggplot() +
  geom_boxplot(aes(x = fct_relevel(harvested, "Harvested", "Unharvested"),
                   y = log(stemden_totaltrees)),
                fill = "mistyrose2", outlier.shape = 21,
                outlier.color = "gray20", outlier.fill = "gray80") +
  plot_theme +
  coord_cartesian(ylim = c(5, 8)) +
  labs(x = "Plot status", y = "Total trees per hectare (log scale)")
ggsave("figures/stemden_fig1.png", height = 4, width = 6)
stemden_fig1
```



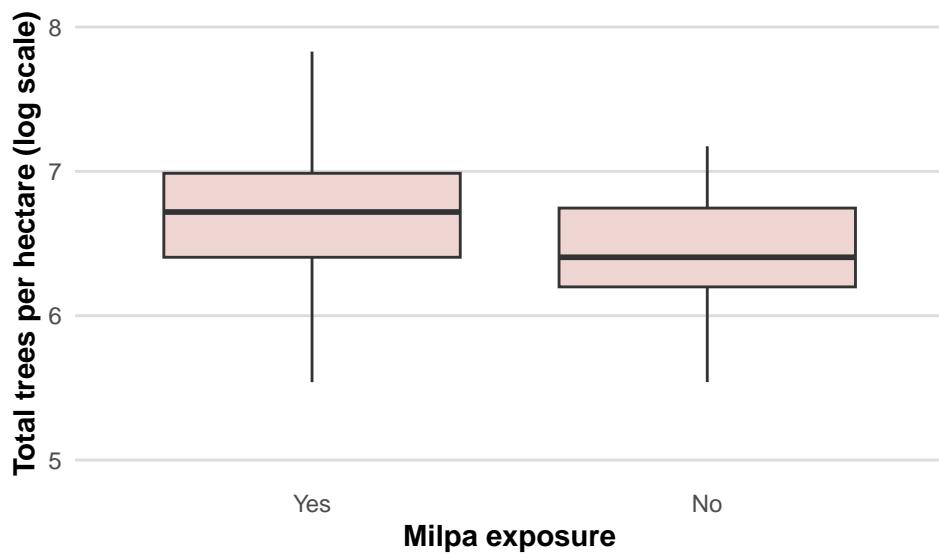
```
# By vegetation type
stemden_fig2 <- data_plots %>%
  mutate(vegetation_type = fct_recode(vegetation_type,
    "Ju'uche'" = "juuche",
    "Keelenche'" = "keelenche",
    "Nuku'uch che'" = "nukuuchche")) %>%
  ggplot() +
  geom_boxplot(aes(x = vegetation_type, y = log(stemden_totaltrees)),
    fill = "mistyrose2", outlier.shape = 21,
    outlier.color = "gray20", outlier.fill = "gray80") +
  plot_theme +
  coord_cartesian(ylim = c(5, 8)) +
  labs(x = "Vegetation type", y = "Total trees per hectare (log scale)")
ggsave("figures/stemden_fig2.png", height = 4, width = 6)
stemden_fig2
```



```

# By milpa exposure
stemden_fig3 <- data_plots %>%
  mutate(milpa = fct_recode(milpa,
                            "Yes" = "yes", "No" = "no")) %>%
  ggplot() +
  geom_boxplot(aes(x = fct_relevel(milpa, "Yes", "No"), y = log(stemden_totaltrees)),
                fill = "mistyrose2", outlier.shape = 21,
                outlier.color = "gray20", outlier.fill = "gray80") +
  plot_theme +
  coord_cartesian(ylim = c(5, 8)) +
  labs(x = "Milpa exposure", y = "Total trees per hectare (log scale)")
ggsave("figures/stemden_fig3.png", height = 4, width = 6)
stemden_fig3

```

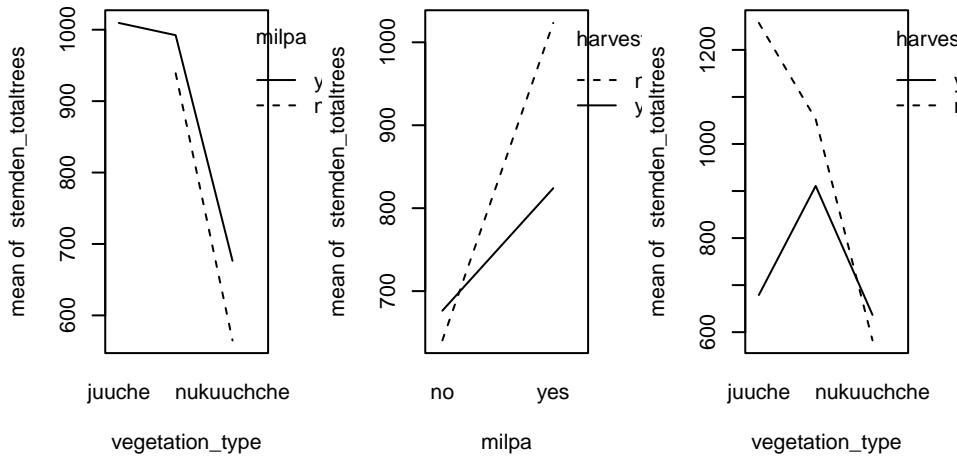


Interaction plots

```

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

```

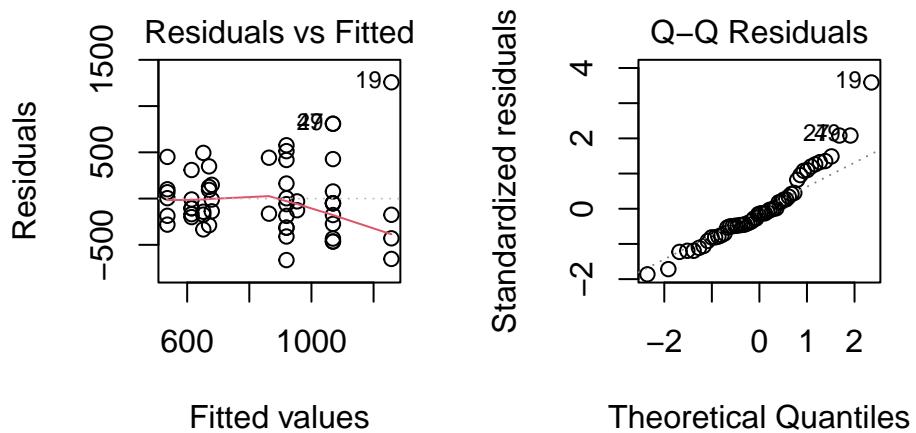


Models

Fit model and check assumptions

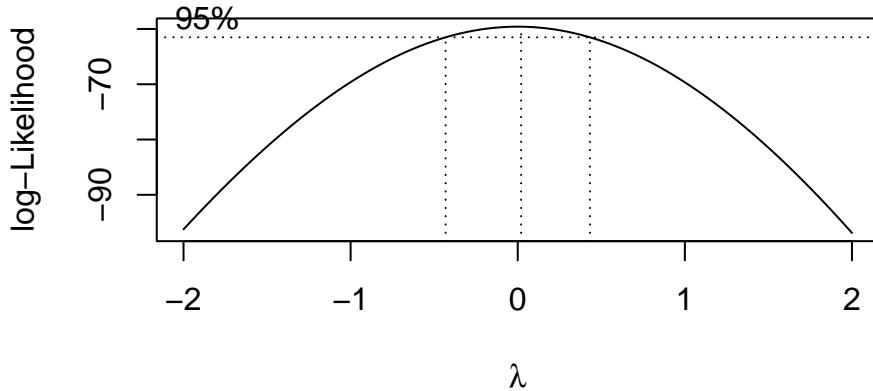
```
mod <- lm(stemden_totaltrees ~ harvested + vegetation_type + milpa +
           harvested:vegetation_type + harvested:milpa, data = data_plots)

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



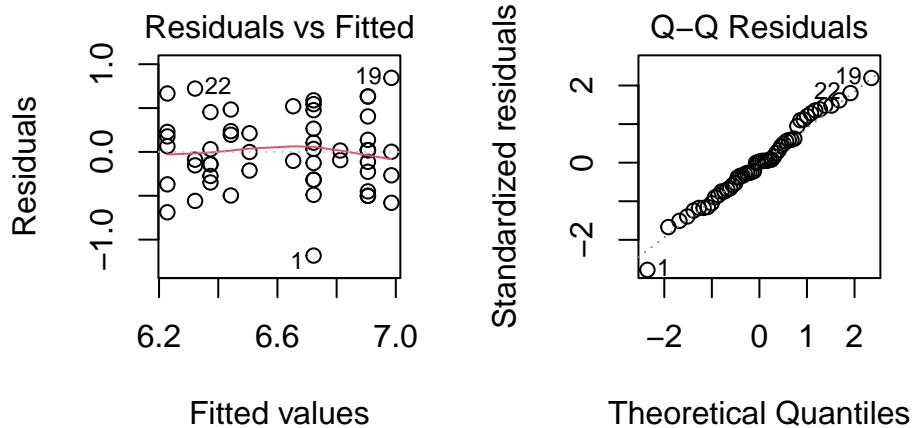
Attempt transformation of response

```
# There is a megaphone shape on the residuals vs. fitted; try Box-Cox
MASS::boxcox(mod)
```



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

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



The assumptions appear to be satisfied now, so we will proceed with the transformed model. Note that we could also fix the assumption violations by removing a significant outlier (plot 19) from the untransformed model. However, we decide to use the log-transform since our sample size is already quite small; if we removed the outlier, it would get even smaller.

Summarize selected model and run ANOVA

```
# Check for outliers
outlierTest(mod.transform)

## No Studentized residuals with Bonferroni p < 0.05
## Largest |rstudent|:
##     rstudent unadjusted p-value Bonferroni p
## 1 -3.011364      0.0042554     0.22979
```

```

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

## 
## Call:
## lm(formula = log(stemden_totaltrees) ~ harvested + vegetation_type +
##     milpa + harvested:vegetation_type + harvested:milpa, data = data_plots)
##
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -1.18231 -0.27230  0.00864  0.23452  0.84486 
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)    
## (Intercept)                6.89154   0.31207 22.083 <2e-16 ***
## harvestedyes              -0.45470   0.45989 -0.989  0.3280    
## vegetation_typekeelenche -0.07930   0.25391 -0.312  0.7562    
## vegetation_typenukuuchche -0.66335   0.29396 -2.257  0.0288 *  
## milpayes                  0.09349   0.21911  0.427  0.6716    
## harvestedyes:vegetation_typekeelenche  0.29577   0.38294  0.772  0.4438  
## harvestedyes:vegetation_typenukuuchche  0.60025   0.43518  1.379  0.1745  
## harvestedyes:milpayes          -0.02461   0.31028 -0.079  0.9371  
## ---                        
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 
##
## Residual standard error: 0.4444 on 46 degrees of freedom
## Multiple R-squared:  0.2785, Adjusted R-squared:  0.1687 
## F-statistic: 2.536 on 7 and 46 DF,  p-value: 0.02705

```

```

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

```

```

## Anova Table (Type II tests)
## 
## Response: log(stemden_totaltrees)
##                               Sum Sq Df F value    Pr(>F)    
## harvested                   0.1366  1  0.6915  0.40996    
## vegetation_type              1.6905  2  4.2794  0.01975 *  
## milpa                      0.0541  1  0.2740  0.60314    
## harvested:vegetation_type   0.4038  2  1.0222  0.36783    
## harvested:milpa              0.0012  1  0.0063  0.93713    
## Residuals                   9.0859 46                        
## ---                        
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 

```

We find that vegetation type is significant, but harvested, milpa, and the interaction terms are not.

Pairwise comparisons

```
# Tukey HSD for vegetation type
# Keelenche' has a significantly higher stem density than Nuku'uch che'
TukeyHSD(aov(mod.transform), "vegetation_type", conf.level = 0.95)
```

```
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = mod.transform)
##
## $vegetation_type
##          diff      lwr      upr     p adj
## keelenche-juuche   0.03171452 -0.4248034 0.48823241 0.9845211
## nukuuchche-juuche -0.43808055 -0.9107604 0.03459932 0.0743329
## nukuuchche-keelenche -0.46979507 -0.7873372 -0.15225298 0.0023133
```

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

```
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = mod.transform)
##
## $harvested
##          diff      lwr      upr     p adj
## yes-no -0.1132648 -0.3569102 0.1303807 0.354291
```