Stem density

Tim White

June 2024

Packages and data

```r
library(car)
library(tidyverse)
library(lme4)
```

```r
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"))
```
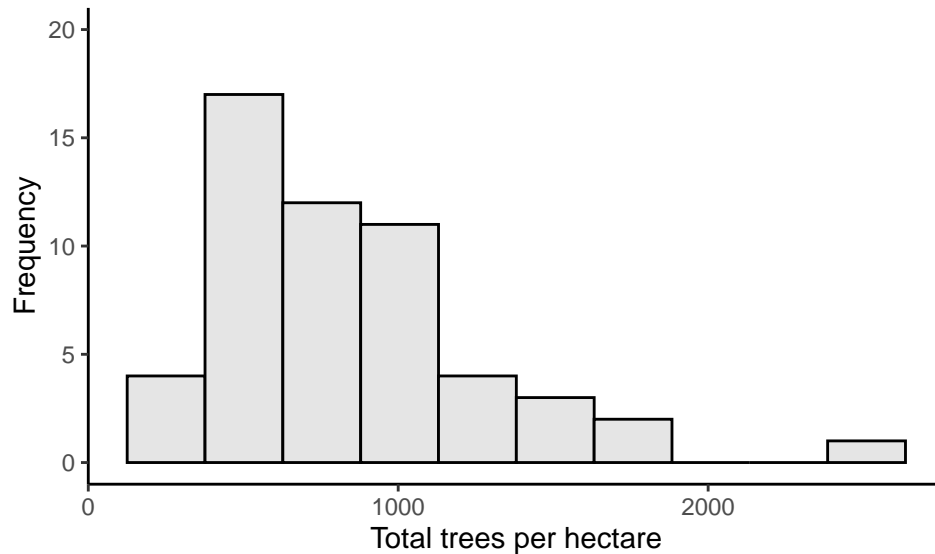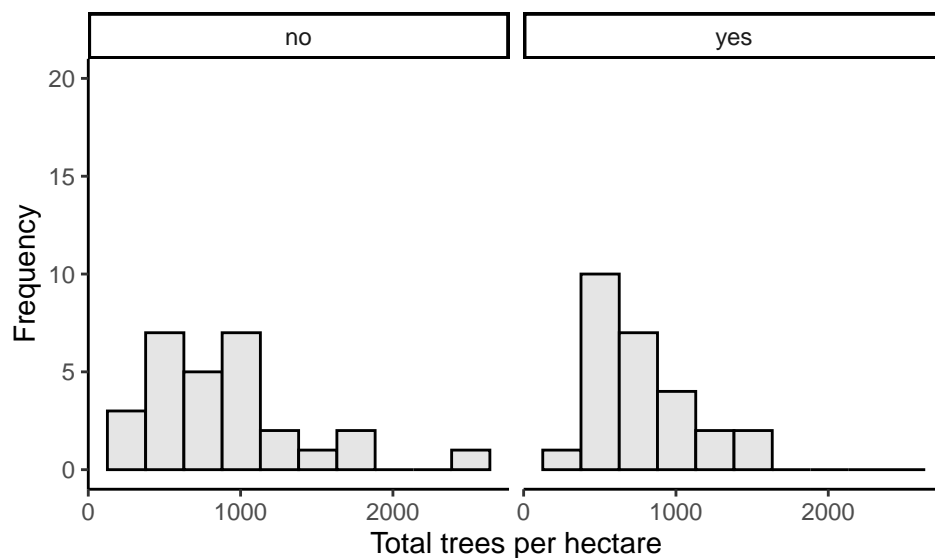
Figures

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

```r
# 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
```

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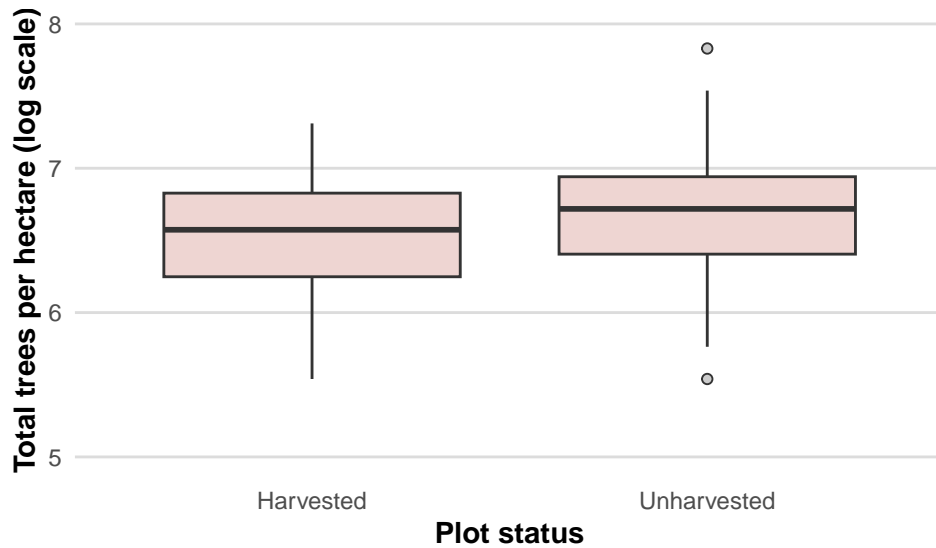
```r
data_plots %>% filter(harvested == "no") %>% pull(stemden_totaltrees) %>% summary()
```

```
##    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)

```r
# 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
```
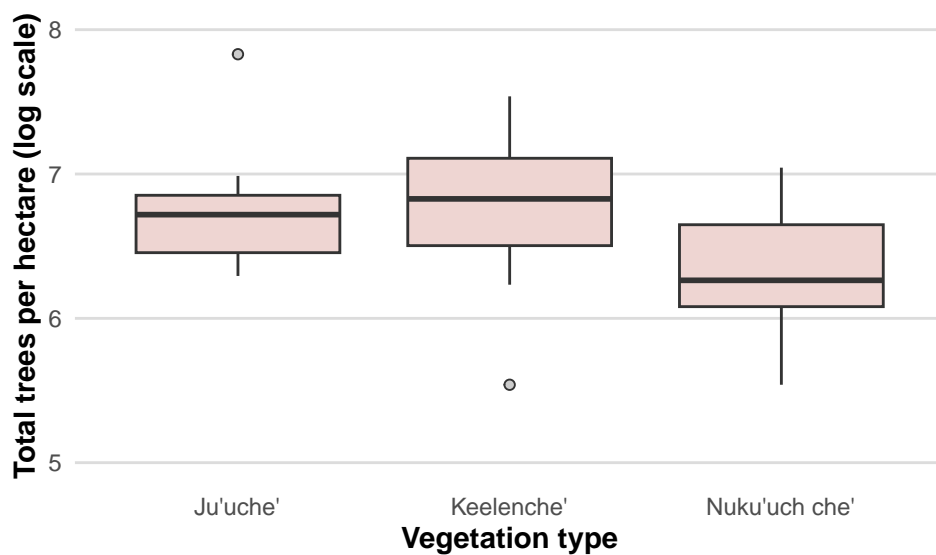
```
# By vegetation type
stemden_fig2 <- data_plots %>%
  mutate(vegetation_type = fct_recode(vegetation_type,
                                      "Ju'uche'" = "juuche",
                                      "Keelenche'" = "keelenche",
                                      "Nuku'uch che'" = "nukuuchche")) %>%
  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
```
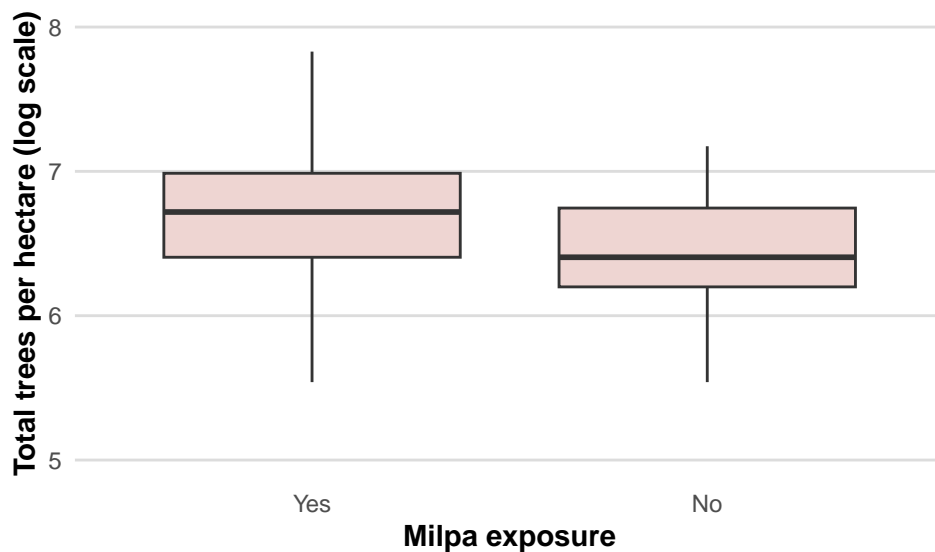
```
# By milpa exposure
stemden_fig3 <- data_plots %>%
  mutate(milpa = fct_recode(milpa,
                            "Yes" = "yes", "No" = "no")) %>%
  ggplot() +
  geom_boxplot(aes(x = fct_relevel(milpa, "Yes", "No"), y = log(stemden_totaltrees)),
               fill = "mistyrose2", outlier.shape = 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
```
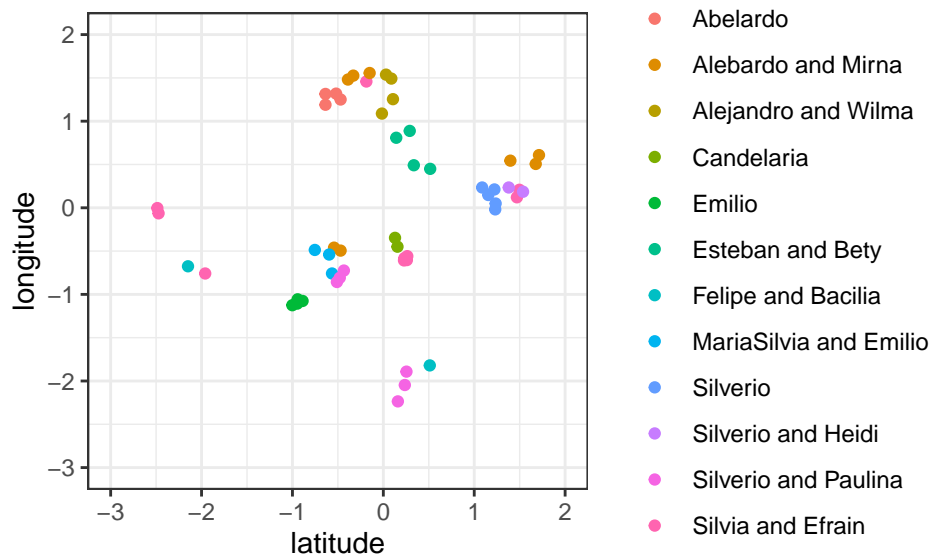

Latitude and longitude by artisan

```
data_plots %>%
  ggplot(aes(x = latitude, y = longitude)) +
    geom_point(aes(color = artisan), shape = 19) +
    xlim(-3, 2) +
    ylim(-3, 2) +
    theme_bw()
```

Models

We have fewer than five observations for each artisan, so we elect not to include artisan as a random effect. Also, some individuals appear in more than one level of artisan, so we could not disentangle variability due to each individual.
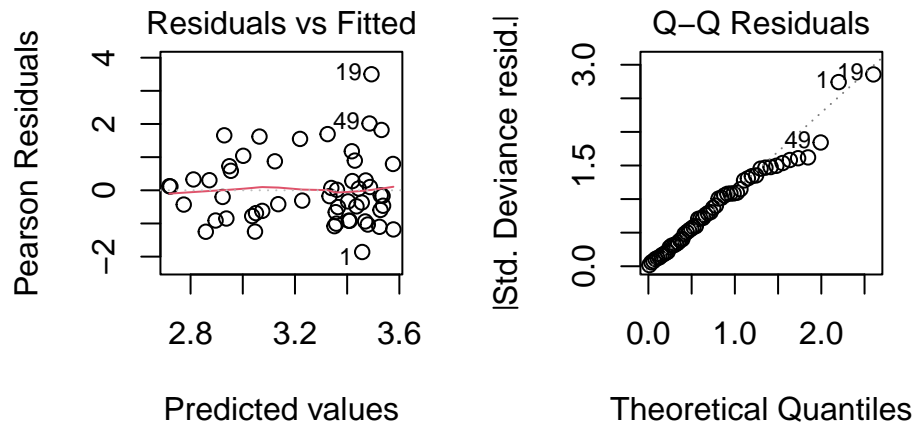
We decide to fit a negative binomial model with all other main effects, including latitude and longitude. We found that the variance of the residuals increases quadratically with the mean, so we believe a negative binomial model is more appropriate than a quasipoisson GLM.

We elect not to include the interaction between harvested and milpa or between harvested and vegetation type. Preliminary analyses found that these interactions were insignificant, and we have no ecological reason to believe that these interactions exist. Removing the interactions also enables us to interpret the main effects of harvested and milpa, which are our primary interests. Removing the interactions also enables a cleaner analysis of deviance table — we don't have to worry about a mismatch between the hypothesis tests run by `summary()` and by `Anova()`.

Fit negative binomial model and check residual plots

```r
mod <- MASS::glm.nb(stemden_totaltrees_count ~ harvested + vegetation_type + milpa +
                    latitude + longitude + offset(log(plots_per_ha)), data = data_plots)
```

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

Summarize the model

```
summary(mod)
```

```
##
## Call:
## MASS::glm.nb(formula = stemden_totaltrees_count ~ harvested +
##     vegetation_type + milpa + latitude + longitude + offset(log(plots_per_ha)),
##     data = data_plots, init.theta = 7.691325011, link = log)
##
## Coefficients:
##                            Estimate Std. Error z value Pr(>|z|)
## (Intercept)                6.738037   0.244187  27.594   <2e-16 ***
## harvestedyes              -0.115548   0.112333  -1.029    0.304
## vegetation_typekeelenche  -0.005475   0.184724  -0.030    0.976
## vegetation_typenukuuchche -0.314026   0.231434  -1.357    0.175
## milpayes                   0.212033   0.165529   1.281    0.200
## latitude                   0.001578   0.061132   0.026    0.979
## longitude                 -0.084035   0.073462  -1.144    0.253
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(7.6913) family taken to be 1)
##
##     Null deviance: 73.688  on 53  degrees of freedom
## Residual deviance: 53.876  on 47  degrees of freedom
## AIC: 418.58
##
## Number of Fisher Scoring iterations: 1
##
##
##               Theta:  7.69
##           Std. Err.:  1.88
##
##  2 x log-likelihood:  -402.58
```

Construct analysis of deviance table

```
# Use type 2 SS since data are unbalanced)
Anova(mod, type = 2, test.statistic = "LR")
```

```
## Analysis of Deviance Table (Type II tests)
##
## Response: stemden_totaltrees_count
##                 LR Chisq Df Pr(>Chisq)
## harvested         1.0407  1     0.3077
## vegetation_type   3.4510  2     0.1781
## milpa             1.6443  1     0.1997
## latitude          0.0007  1     0.9788
## longitude         1.3218  1     0.2503
```