

Tree plantations inserted in a benign environmental context can contribute to the maintenance of forest avifauna

Rayssa T. Amarante Ivana Cardoso Raquel T. Amarante
Felipe Brum Anderson S. Bueno

2025-06-03

Table of contents

1 Introduction	5
1.1 Data Access	6
1.2 Attribution	6
2 Characteristics of the tree plantations	7
Ensuring reproducibility	7
Loading packages	7
Importing data	7
Preparing data	7
2.1 Age	8
2.1.1 Performing a t-test	8
2.1.2 Plotting the graph	9
2.2 Distance to forest	10
2.2.1 Performing a t-test	10
2.2.2 Plotting the graph	11
2.3 Understory vegetation obstruction	13
2.3.1 Fitting an ANOVA model	13
2.3.2 Plotting the graph	14
2.3.3 Saving the graph	16
3 Indicator Species	17
Ensuring reproducibility	17
Loading packages	17
Importing data	17
Preparing data	17
3.1 Identifying indicator species	18
3.1.1 Calculates the indicator value for each species	18
3.1.2 Summary the results	18
4 Bird Species Richness	22
Ensuring reproducibility	22
Loading packages	22
Importing data	22
Preparing data	23

4.1	Forest type	23
4.1.1	Preparing data	23
4.1.2	Creating rarefaction curves	23
4.1.3	Plotting the graph	23
4.1.4	Saving the graph	25
4.2	Understorey vegetation obstruction	25
4.2.1	Preparing data	25
4.2.2	Fitting the linear regression model	26
4.2.3	Plotting the graph	27
4.2.4	Saving the graph	27
5	Bird Species Abundance	29
	Ensuring Reproducibility	29
	Load packages	29
	Importing data	29
	Preparing data	30
5.1	Forest type	30
5.1.1	Preparing data	30
5.1.2	Calculating ANOVA	30
5.1.3	Testing normality of ANOVA residuals	30
5.1.4	Calculating Kruskal-Wallis test	31
5.1.5	Calculating Wilcoxon test	31
5.1.6	Plotting the graph	32
5.1.7	Saving the graph	32
5.2	Understory vegetation obstruction	33
5.2.1	Preparing data	33
5.2.2	Fitting the regression model	33
5.2.3	Plotting the graph	34
5.2.4	Saving the graph	35
5.3	Indicator species	36
5.3.1	Preparing data	36
5.3.2	Plotting the graph	37
5.3.3	Saving the graph	38
6	Bird Community Integrity	39
	Ensuring Reproducibility	39
	Load packages	39
	Importing data	39
	Preparing data	39
6.1	Forest type	40
6.1.1	Calculating Community Integrity	40
6.1.2	Calculating ANOVA	41
6.1.3	Plotting the graph	42

6.1.4	Saving the graph	43
6.2	Understory vegetation obstruction	43
6.2.1	Preparing data	43
6.2.2	Fitting the regression model	44
6.2.3	Plotting the graph	45
6.2.4	Saving the graph	45

1 Introduction

This document covers the analyses used in the paper *Tree plantations inserted in a benign environmental context can contribute to the maintenance of forest avifauna*, published in the journal **Forest Ecology and Management**.

The objectives of this study were:

- (1) Investigate the effects of different forest types (Fig. 1) embedded within a benign environmental context - characterised by old stand age, the presence of understorey vegetation, and proximity to natural forest patches - on species richness, abundance and composition of understorey birds
- (2) Determine potential indicator species for each forest type.



Figure 1.1: Fig. 1 The three forest types sampled in the Floresta Nacional de Passo Fundo, southern Brazil: natural forest (Mixed Ombrophilous Forest), araucaria plantation (native species), and pine plantation (exotic species)

This document is organised into sections, which you can navigate using the links on the left sidebar.

1.1 Data Access

All data used in the paper and in this code are available on the KNB platform: <https://knb.ecoinformatics.org/view/doi:10.5063/F1QR4VK1>

1.2 Attribution

If you need any clarification or are interested in the project, please contact:

Ivana Cardoso (ivanawaters@gmail.com)
PhD student, Instituto Nacional de Pesquisas da Amazônia

2 Characteristics of the tree plantations

We assessed whether the native (*Araucaria angustifolia*) and exotic (*Pinus elliottii*) tree plantations differ in three aspects: age, understorey vegetation obstruction, and distance to the nearest natural forest patch larger than 10 hectares.

Ensuring reproducibility

```
set.seed(26)
```

Loading packages

```
library(stats)
library(ggplot2)
library(ggpubr)
library(patchwork)
library(rstatix)
```

Importing data

```
# https://knb.ecoinformatics.org/view/doi:10.5063/F1QR4VK1
setwd("C:/Users/ivana/OneDrive/FLONA_aves/data")
data <- read.csv("FLONA-PF_dataset.txt", sep = "\t")
env_all <- data[,1:8]
env_plant <- subset(env_all, forest.type != "Natural_forest")
```

Preparing data

```
color <- rep(c("#912DBF", "#BF912D"), each = 6)
color_complete <- rep(c("#2DBF91", "#912DBF", "#BF912D"), each = 6)
```

2.1 Age

2.1.1 Performing a t-test

```
# Perform the Shapiro-Wilk test to check if age data is normally distributed
shapiro.test(env_plant$age) # It's normal (p > 0.05)
```

```
Shapiro-Wilk normality test

data: env_plant$age
W = 0.94715, p-value = 0.5957

# Perform Levene's test to check for homogeneity of variances across forest types
levene_test(age ~ forest.type, data = env_plant) # The variances are homogeneous (p > 0.05)

Warning in leveneTest.default(y = y, group = group, ...): group coerced to
factor.

# A tibble: 1 x 4
  df1   df2 statistic     p
<int> <int>    <dbl> <dbl>
1     1     10      0.546 0.477

# Perform a t-test to compare age between forest types
t.test(env_plant$age ~ env_plant$forest.type, var.equal = TRUE)
```

Two Sample t-test

```
data: env_plant$age by env_plant$forest.type
t = 4.6696, df = 10, p-value = 0.0008813
alternative hypothesis: true difference in means between group Araucaria_plantation and group
```

```

95 percent confidence interval:
 4.967017 14.032983
sample estimates:
mean in group Araucaria_plantation      mean in group Pine_plantation
          62.66667                          53.16667

```

There was a significant difference in age between the two plantation types ($t = 4.6696$, $p < 0.001$).

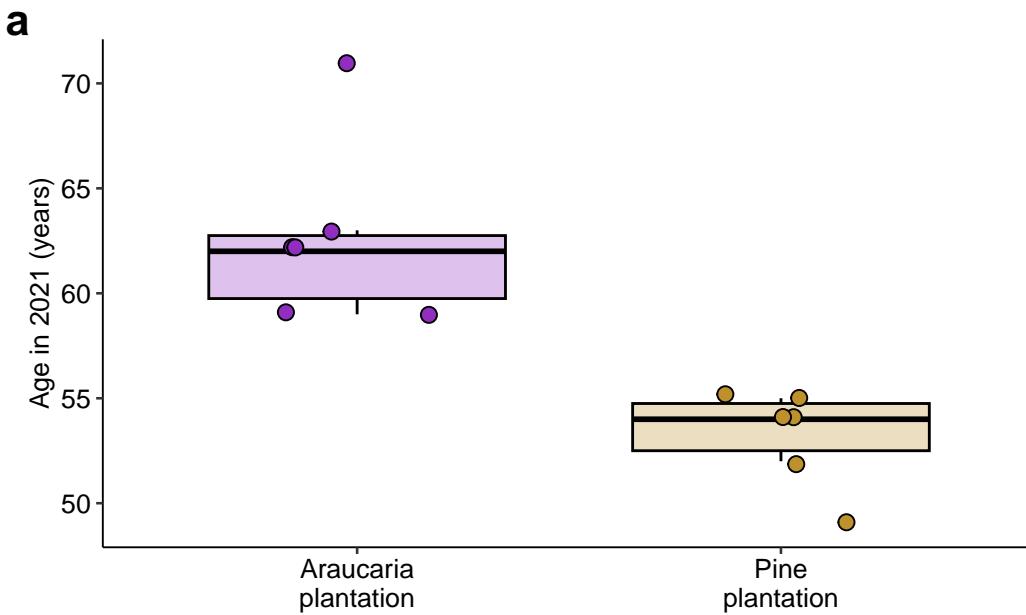
2.1.2 Plotting the graph

```

plot_age <-
  ggplot(data = env_plant,
         aes(x = forest.type, y = age)) +
  scale_y_continuous(breaks = c(50, 55, 60, 65, 70, 75)) +
  geom_boxplot(outlier.shape = NA,
               width = 0.7, col = "black", fill = c("#912DBF", "#BF912D"),
               alpha = 0.30, show.legend = FALSE) +
  geom_jitter(shape= 21,
              width = 0.2, height = 0.2, colour = "black",
              fill = color, size = 2.5) +
  scale_x_discrete(labels = c("Araucaria\nplantation",
                             "Pine\nplantation")) +
  labs(x = " ", y = "Age in 2021 (years)", colour="black",
       tag = "a") +
  theme_pubr (base_size = 10) +
  theme(axis.line = element_line(linewidth = 1/3))+ 
  theme(plot.tag = element_text(size = 16, face = "bold"))

plot_age

```



2.2 Distance to forest

2.2.1 Performing a t-test

```
# Perform the Shapiro-Wilk test to check if the distances to the nearest natural forest are normal
shapiro.test(env_plant$distance.forest.m) # It's normal (p > 0.05)
```

```
Shapiro-Wilk normality test

data: env_plant$distance.forest.m
W = 0.89872, p-value = 0.1527

# Perform Levene's test to check for homogeneity of variances across forest types
levene_test(distance.forest.m ~ forest.type, data = env_plant) # The variances are homogeneous
```

Warning in leveneTest.default(y = y, group = group, ...): group coerced to factor.

```

# A tibble: 1 x 4
  df1   df2 statistic     p
  <int> <int>     <dbl> <dbl>
1     1     10  0.000242 0.988

# Perform a t-test to compare the distances to the nearest natural forest between forest type
t.test(env_plant$distance.forest.m ~ env_plant$forest.type, var.equal = TRUE)

```

Two Sample t-test

```

data: env_plant$distance.forest.m by env_plant$forest.type
t = -0.12128, df = 10, p-value = 0.9059
alternative hypothesis: true difference in means between group Araucaria_plantation and group
95 percent confidence interval:
-749.5126 672.1339
sample estimates:
mean in group Araucaria_plantation      mean in group Pine_plantation
                                         764.7762                         803.4655

```

The distance of the tree plantations to the nearest natural forest did not differ significantly between the two types of plantation ($t = -0.121$; $p = 0.906$).

2.2.2 Plotting the graph

```

plot_dist <-
  ggplot(data = env_plant,
         aes(x = forest.type, y = distance.forest.m)) +
  geom_boxplot(outlier.shape = NA,
               width = 0.7, col = "black",
               fill = c("#912DBF", "#BF912D"),
               alpha = 0.30, show.legend = FALSE) +
  geom_jitter(shape= 21, width = 0.2,
              height = 0.2, colour = "black",
              fill = color, size = 2.5) +
  scale_x_discrete(labels = c("Araucaria\\nplantation",

```

```

  "Pine\ plantation")) +  

  scale_y_continuous(breaks = c(200, 600, 1000, 1400, 1800)) +  

  labs(x = " ",  

       y = "Distance to natural forest (m)",  

       colour="black",  

       tag = "b") +  

  theme_pubr (base_size = 10) +  

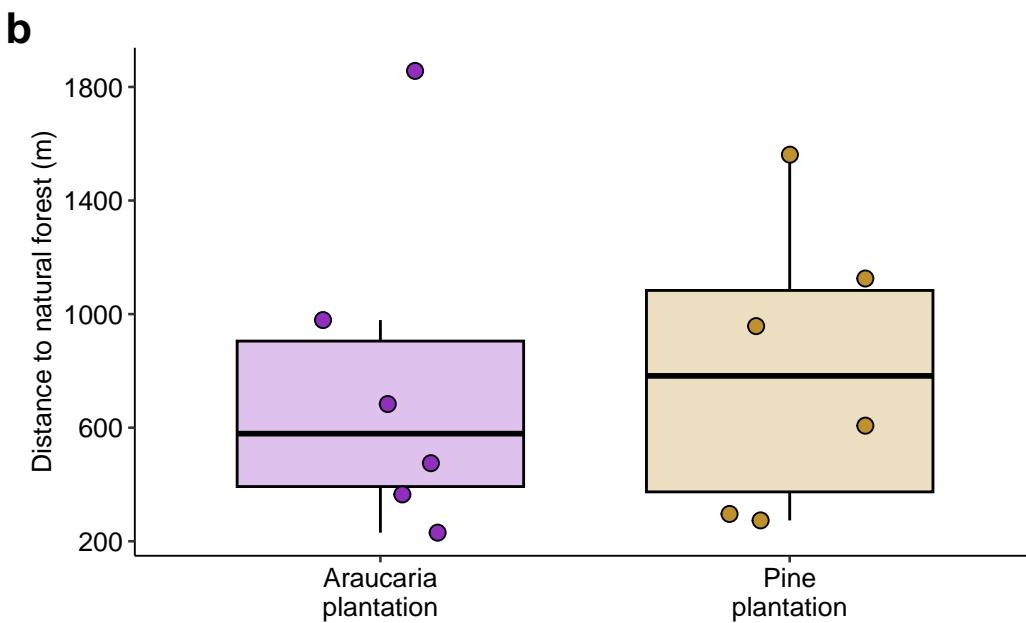
  theme(axis.line = element_line(linewidth = 1/3)) +  

  theme(plot.tag = element_text(size = 16, face = "bold"))

plot_dist

```



2.3 Understory vegetation obstruction

2.3.1 Fitting an ANOVA model

```
# Perform the Shapiro-Wilk test to check if understorey vegetation obstruction is normally distributed
shapiro.test(env_all$obstruction) # It's normal (p > 0.05)
```

```
Shapiro-Wilk normality test

data: env_all$obstruction
W = 0.94235, p-value = 0.3177
```

```
# Perform Levene's test to check for homogeneity of variances across forest types
levene_test(obstruction ~ as.factor(forest.type), data = env_all) # The variances are homogeneous
```

```
# A tibble: 1 x 4
  df1   df2 statistic     p
  <int> <int>    <dbl> <dbl>
1     2     15     1.22 0.322
```

```
# Perform one-way ANOVA to test for differences in understorey vegetation obstruction between forest types
anova <- aov(obstruction ~ forest.type, data = env_all)
summary(anova)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)						
forest.type	2	630.8	315.39	5.986	0.0123 *						
Residuals	15	790.3	52.69								

Signif. codes:	0	'***'	0.001	'**'	0.01	'*'	0.05	'.'	0.1	' '	1

```
# Perform the Shapiro-Wilk test on the residuals to check if they are normally distributed
shapiro.test(anova$residuals) # Residuals of the ANOVA model are normal (p > 0.05)
```

```
Shapiro-Wilk normality test

data: anova$residuals
W = 0.9706, p-value = 0.8089
```

```

# Perform Levene's test on residuals to check for homogeneity of variances
levene_test(anova$residuals ~ as.factor(forest.type), data = env_all) # The variances are homogenous

# A tibble: 1 x 4
  df1   df2 statistic     p
  <int> <int>    <dbl> <dbl>
1     2     15      1.22  0.322

# Perform Tukey's HSD post-hoc test to compare pairwise differences between forest types
par_par <- TukeyHSD(anova)
par_par

Tukey multiple comparisons of means
95% family-wise confidence level

Fit: aov(formula = obstruction ~ forest.type, data = env_all)

$forest.type
            diff      lwr      upr      p adj
Natural_forest-Araucaria_plantation  2.166667 -8.718861 13.052194 0.8642904
Pine_plantation-Araucaria_plantation -11.333333 -22.218861 -0.447806 0.0408047
Pine_plantation-Natural_forest       -13.500000 -24.385527 -2.614473 0.0148458

```

Understorey vegetation obstruction significantly differed between forest types ($F = 5.986$, $p = 0.01$). We found significant differences between pine plantations and natural forests (diff = -13.50, $p = 0.015$) and between pine plantations and araucaria plantations (diff = -11.33, $p = 0.041$). No significant difference was found between araucaria plantations and natural forests (diff = 2.16, $p = 0.864$).

2.3.2 Plotting the graph

```

# Renaming forest types for consistent ordering in the plot
env_all$forest.type[env_all$forest.type == "Natural_forest"] <- "1.Natural_forest"

env_all$forest.type[env_all$forest.type == "Araucaria_plantation"] <- "2.Araucaria_plantation"

env_all$forest.type[env_all$forest.type == "Pine_plantation"] <- "3.Pine_plantation"

plot_obs <-

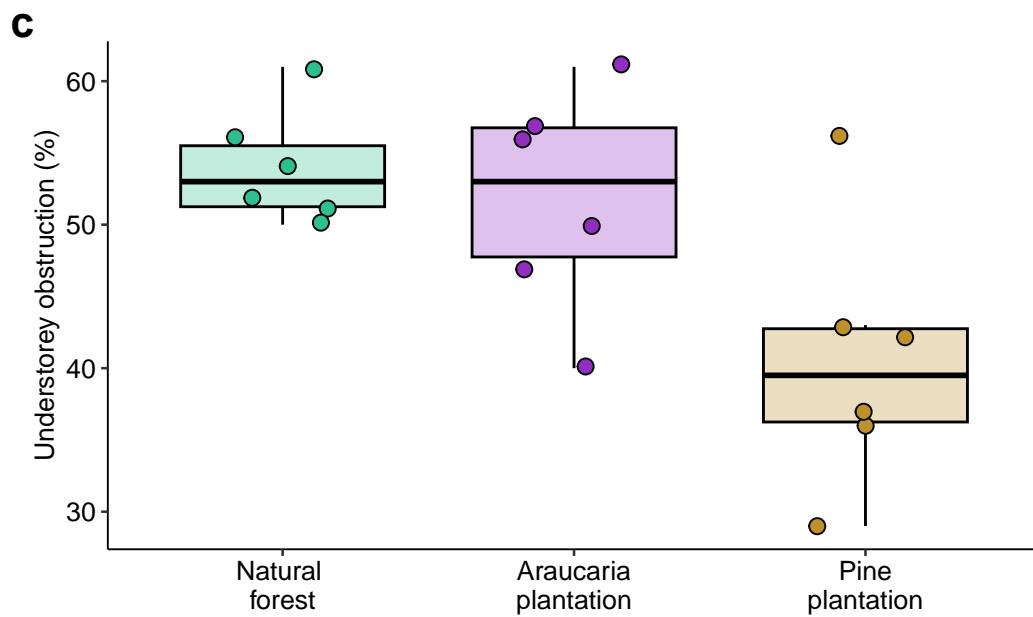
```

```

ggplot(data = env_all,
       aes(x = forest.type, y = obstruction)) +
  geom_boxplot(outlier.shape = NA,
               width = 0.7, col = "black",
               fill = c("#2DBF91", "#912DBF", "#BF912D"),
               alpha = 0.30, show.legend = FALSE) +
  geom_jitter(shape= 21,
              width = 0.2, height = 0.2, colour = "black",
              fill = color_complete, size = 2.5) +
  scale_x_discrete(labels = c("Natural \nforest",
                             "Araucaria\nplantation",
                             "Pine\nplantation")) +
  labs(x = " ", y = "Understorey obstruction (%)", colour="black",
       tag = "c") +
  theme_pubr (base_size = 10) +
  theme(axis.line = element_line(linewidth = 1/3)) +
  theme(plot.tag = element_text(size = 16, face = "bold"))

plot_obs

```



2.3.3 Saving the graph

```
ggsave(
  plot = plot_age + plot_dist + plot_obs,
  filename = "charact_plantation.png", dpi = 500,
  width = 7 * 3, height = 7, units = 'cm')
```

3 Indicator Species

To identify indicator species associated with each forest type, we conducted an Indicator Species Analysis using the *multipatt* function from the *indicspecies* package. The threshold for the IndVal index was set at 0.70.

Ensuring reproducibility

```
set.seed(1341562062)
```

Loading packages

```
library(indicspecies)
library(vegan)
```

Importing data

```
# https://knb.ecoinformatics.org/view/doi:10.5063/F1QR4VK1
setwd("C:/Users/ivana/OneDrive/FLONA_aves/data")
data <- read.csv("FLONA-PF_dataset.txt", sep = "\t")
```

Preparing data

```
grup <- as.factor(data$forest.type)
```

3.1 Identifying indicator species

3.1.1 Calculates the indicator value for each species

```
indval <- multipatt(data[9:59], cluster = grup,
                      func = "IndVal.g", control = how(nperm = 999))

indval <- as.vector(indval)
```

3.1.2 Summary the results

```
summary(indval, indvalcomp = TRUE)
```

```
Multilevel pattern analysis
-----
Association function: IndVal.g
Significance level (alpha): 0.05

Total number of species: 51
Selected number of species: 5
Number of species associated to 1 group: 1
Number of species associated to 2 groups: 4

List of species associated to each combination:

Group Natural_forest #sps. 1
          A   B stat p.value
Leptopogon_amaurocephalus 0.8 1.0 0.894  0.001 ***
                                         A   B stat p.value
Group Araucaria_plantation+Natural_forest #sps. 2
                                         A   B stat p.value
Synallaxis_cinerascens 0.8723 1.0000 0.934  0.005 **
Chiropipia_caudata    0.9365 0.9167 0.927  0.020 *
                                         A   B stat p.value
Group Araucaria_plantation+Pine_plantation #sps. 1
                                         A   B stat p.value
Zonotrichia_capensis 0.9107 1.0000 0.954  0.007 **
```

```
Group Natural_forest+Pine_plantation #sps. 1
      A      B  stat p.value
Turdus_leucomelas 1.0000 0.6667 0.816   0.031 *
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```



Figure 3.1: Fig. 1 *Leptopogon amaurocephalus*, a species associated with **natural forests**.



Figure 3.2: Fig. 2 *Synallaxis cinerascens*, a species associated with **araucaria plantations and natural forests**.



Figure 3.3: Fig. 3 *Chiroxiphia caudata*, a species associated with **araucaria plantations and natural forests**.



Figure 3.4: Fig. 4 *Zonotrichia capensis*, a species associated with **araucaria and pine plantations**.



Figure 3.5: Fig. 5 *Turdus leucomelas*, a species associated with **pine plantations and natural forests**.

4 Bird Species Richness

Predictor variables:

1. **Forest type** - We compared species richness among the three forest types using rarefaction curves from *iNEXT* package. These curves were based on 1000 bootstrap replications. We considered differences in species richness to be significant when the 95% confidence intervals of the curves did not overlap.
2. **Understorey vegetation obstruction** - We performed simple linear regression analyses to investigate the effect of understorey vegetation obstruction on species richness.

Ensuring reproducibility

```
set.seed(26)
```

Loading packages

```
library(ggplot2)
library(iNEXT)
library(ggpubr)
library(dplyr)
library(rstatix)
```

Importing data

```
# https://knb.ecoinformatics.org/view/doi:10.5063/F1QR4VK1
setwd("C:/Users/ivana/OneDrive/FLONA_aves/data")
data <- read.csv("FLONA-PF_dataset.txt", sep = "\t")
birds <- data[,c(9:59)]
```

Preparing data

```
# Renaming forest types for consistent ordering in the plot  
data$forest.type <- recode(data$forest.type,  
                           "Natural_forest" = "1.Natural_forest",  
                           "Araucaria_plantation" = "2.Araucaria_plantation",  
                           "Pine_plantation" = "3.Pine_plantation")
```

4.1 Forest type

4.1.1 Preparing data

```
# Group the rows of the species matrix by habitat and sums the species values within each habitat  
raref = aggregate(birds, by = list(habitat = data[1:18,2]), FUN = sum)  
  
# Set the row names  
rownames(raref) = raref$habitat  
  
# Remove the first column, keeping only the species count columns  
raref = raref[-1]  
  
# Transpose the data frame  
raref = as.data.frame(t(raref))
```

4.1.2 Creating rarefaction curves

```
# Estimates rarefaction and extrapolation curves of species richness (q = 0) based on abundance  
raref <- iNEXT(raref, q= 0, datatype = "abundance", endpoint = NULL, conf = 0.95, nboot = 1000)
```

4.1.3 Plotting the graph

```
# Extracts the size-based estimation data from the iNEXT results.  
y = raref$iNextEst$size_based  
  
# Removes extrapolated estimates
```

```

z = y[which(y$Method != "Extrapolation"),]

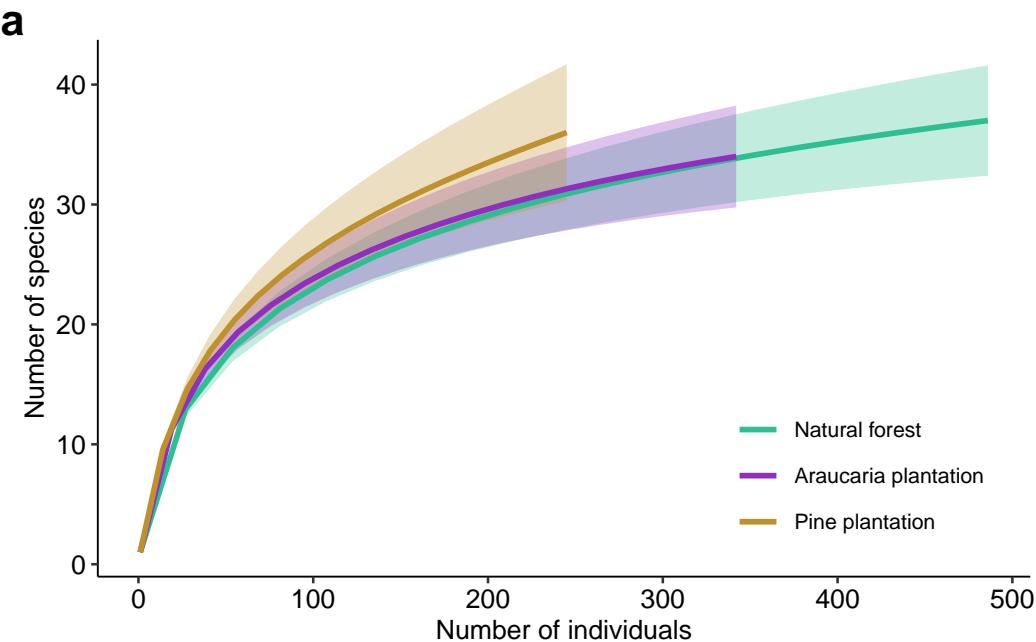
# Converts the 'Assemblage' column to a factor with levels ordered by their appearance in the
z$Assemblage = factor(z$Assemblage,
                       levels = unique(z$Assemblage))

plot_rich1 <-
  ggplot(data = z,
         aes(x = m, y = qD,
             colour = Assemblage,
             fill = Assemblage)) +
  labs(y = "Number of species",
       x = "Number of individuals",
       colour = NULL, fill = NULL,
       tag = "a") +
  scale_color_manual(values = c("#2DBF91", "#912DBF", "#BF912D"),
                     labels = c("Natural forest",
                               "Araucaria plantation",
                               "Pine plantation")) +
  scale_fill_manual(values = c("#2DBF91", "#912DBF", "#BF912D"),
                    labels = c("Natural forest",
                               "Araucaria plantation",
                               "Pine plantation")) +
  geom_ribbon(aes(ymin = qD.LCL, ymax = qD.UCL),
              colour = NA, alpha = 0.30,
              show.legend = FALSE) +
  geom_line(size = 1) +
  theme_pubr(base_size = 10) +
  theme(legend.position = c(0.95, 0.05),
        legend.justification = c(0.95, 0.05),
        axis.line = element_line(linewidth = 1/3)) +
  theme(plot.tag = element_text(size = 16, face = "bold"))

```

4.1.4 Saving the graph

```
plot_rich1
```



```
ggsave(  
  plot = plot_rich1,  
  filename = "richness_foresttype.png", dpi = 500,  
  width = 7, height = 7, units = 'cm')
```

4.2 Understorey vegetation obstruction

4.2.1 Preparing data

```
species <- data[,c(1:2, 9:59)]  
  
# Calculates species richness for each sample  
species["richness"] <- rowSums(birds>0)  
  
# Creates a new data frame with forest type, species richness, and understorey vegetation ob...
```

```
set <- data.frame(forest.type = data$forest.type,
                   richness = species$richness,
                   obstruction = data$obstruction)
```

4.2.2 Fitting the linear regression model

```
# Fits the model to assess the effect of understorey vegetation obstruction on species richness
reg_rich <- lm(richness ~ obstruction, data = set)
summary(reg_rich)
```

Call:

```
lm(formula = richness ~ obstruction, data = set)
```

Residuals:

Min	1Q	Median	3Q	Max
-6.0453	-2.1640	-0.1846	2.6255	5.0749

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	14.67834	4.47391	3.281	0.00471 **
obstruction	0.06013	0.09024	0.666	0.51470

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.402 on 16 degrees of freedom
Multiple R-squared: 0.027, Adjusted R-squared: -0.03381
F-statistic: 0.444 on 1 and 16 DF, p-value: 0.5147

```
# Tests the normality of model residuals
shapiro.test(reg_rich$residuals) # It's normal (p > 0.05)
```

Shapiro-Wilk normality test

```
data: reg_rich$residuals
W = 0.95376, p-value = 0.487
```

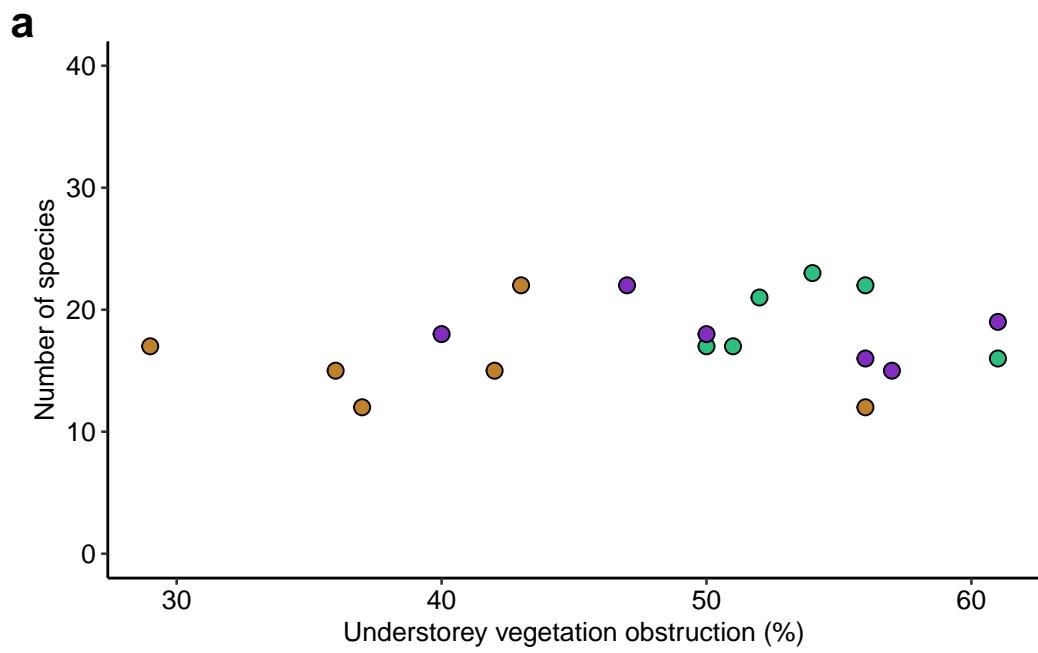
Understorey vegetation obstruction had no significant effect on species richness ($p = 0.5147$).

4.2.3 Plotting the graph

```
plot_rich2 <-
  ggplot(set,
         aes(x = obstruction, y = richness,
             fill = forest.type)) +
  geom_point(colour = "black", size = 2.5,
             shape = 21) +
  scale_y_continuous(limits = c(0, 40),
                     breaks = c(0,10,20,30, 40)) +
  scale_fill_manual(values = c("#2DBF81", "#812DBF", "#BF812D"),
                    labels = c("Natural forest",
                              "Araucaria plantation",
                              "Pine plantation")) +
  labs(x = "Understorey vegetation obstruction (%)",
       y = "Number of species",
       tag = "a") +
  theme_pubr(base_size = 10) +
  theme(legend.position = "none") +
  theme(plot.tag = element_text(size = 16, face = "bold"))
```

4.2.4 Saving the graph

```
plot_rich2
```



```
ggsave(  
  plot = plot_rich2,  
  filename = "richness_understory.png", dpi = 500,  
  width = 7, height = 7, units = 'cm')
```

5 Bird Species Abundance

Predictor variables:

1. **Forest type** - We used the Kruskal-Wallis test to assess the effect of forest type on species abundances. We then applied the pairwise Wilcoxon test to identify significant differences between forest types.
2. **Understorey vegetation obstruction** - We performed simple linear regression analyses to investigate the effect of understorey vegetation obstruction on species abundance.

Ensuring Reproducibility

```
set.seed(26)
```

Load packages

```
library(ggplot2)
library(ggpubr)
library(dplyr)
```

Importing data

```
# https://knb.ecoinformatics.org/view/doi:10.5063/F1QR4VK1
setwd("C:/Users/ivana/OneDrive/FLONA_aves/data")
data <- read.csv("FLONA-PF_dataset.txt", sep = "\t")
abund <- data[,c(1:2, 9:59)]
env_all <- data[,1:8]
```

Preparing data

```
# Renaming forest types for consistent ordering in the plot  
abund$forest.type <- recode(abund$forest.type,  
                           "Natural_forest" = "1.Natural_forest",  
                           "Araucaria_plantation" = "2.Araucaria_plantation",  
                           "Pine_plantation" = "3.Pine_plantation")
```

5.1 Forest type

5.1.1 Preparing data

```
# Calculates species abundance per sample  
abund["abundance"] <- rowSums(data[9:59])
```

5.1.2 Calculating ANOVA

```
# Performs one-way ANOVA to test for differences in species abundance among forest types  
anova <- aov(abundance ~ forest.type, data = abund)
```

5.1.3 Testing normality of ANOVA residuals

```
# Tests whether ANOVA residuals are normally distributed  
shapiro.test(anova$residuals) # It's not normal (p < 0.05)
```

```
Shapiro-Wilk normality test  
  
data: anova$residuals  
W = 0.88687, p-value = 0.03412
```

Since the ANOVA residuals were not normally distributed, we used the Kruskal–Wallis test instead. To compare groups, we then ran pairwise Wilcoxon tests.

5.1.4 Calculating Kruskal-Wallis test

```
# Performs the Kruskal-Wallis test to compare species abundance among forest types
kruskal <- kruskal.test(abundance ~ forest.type, data = abund)
kruskal
```

```
Kruskal-Wallis rank sum test

data: abundance by forest.type
Kruskal-Wallis chi-squared = 10.561, df = 2, p-value = 0.005089
```

The number of individuals varied significantly between forest types (Kruskal-Wallis test: $\chi^2 = 10.561$, $df = 2$, $p = 0.005$).

5.1.5 Calculating Wilcoxon test

```
# Performs pairwise Wilcoxon tests to compare forest types, with p-value adjustment using the BH method
pair <- pairwise.wilcox.test(abund$abundance, abund$forest.type, p.adjust.method = "BH")
pair
```

```
Pairwise comparisons using Wilcoxon rank sum exact test

data: abund$abundance and abund$forest.type

  1.Natural_forest 2.Araucaria_plantation
2.Araucaria_plantation 0.023          -
3.Pine_plantation    0.023          0.041

P value adjustment method: BH
```

Wilcoxon pairwise comparisons showed that species were more abundant in natural forests than in araucaria plantations ($p = 0.023$) and pine plantations ($p = 0.023$). There was also a difference in species abundance between the two plantation types ($p = 0.041$).

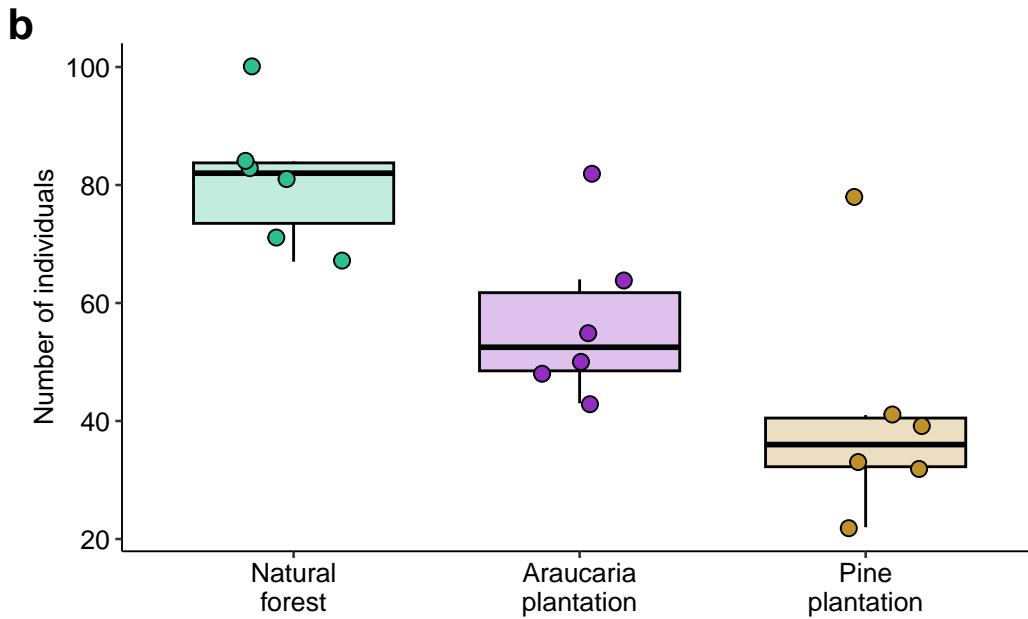
5.1.6 Plotting the graph

```
# Creates a vector of colours, assigning one colour to each forest type (6 samples per type)
color <- rep(c("#2DBF91", "#912DBF", "#BF912D"), each = 6)

plot_abund1 <-
  ggplot(data = abund,
         aes(x = forest.type, y = abundance)) +
  
  geom_boxplot(outlier.shape = NA,
               width = 0.7, col = "black",
               fill = c("#2DBF91", "#912DBF", "#BF912D"),
               alpha = 0.30, show.legend = FALSE) +
  
  geom_jitter(shape = 21, width = 0.2,
              height = 0.2, colour = "black",
              fill = color, size = 2.5) +
  
  scale_x_discrete(labels = c("Natural\nforest",
                               "Araucaria\nplantation",
                               "Pine\nplantation"))+
  
  labs(x = " ", y = "Number of individuals",
       colour="black",
       tag = "b") +
  
  theme_pubr (base_size = 10) +
  
  theme(axis.line = element_line(linewidth = 1/3)) +
  
  theme(plot.tag = element_text(size = 16, face = "bold"))
```

5.1.7 Saving the graph

```
plot_abund1
```



```
ggsave(
  plot = plot_abund1,
  filename = "abundance_foresttype.png", dpi = 500,
  width = 7, height = 7, units = 'cm')
```

5.2 Understory vegetation obstruction

5.2.1 Preparing data

```
# Creates a new data frame with forest type, species abundance, and understory vegetation obstruction
set <- data.frame(forest.type = env_all$forest.type,
                   abundance = abund$abundance,
                   obstruction = env_all$obstruction)
```

5.2.2 Fitting the regression model

```
# Fits a linear model to assess the effect of vegetation obstruction on species abundance
reg_abund <- lm(set$abundance ~ set$obstruction)
summary(reg_abund)
```

```

Call:
lm(formula = set$abundance ~ set$obstruction)

Residuals:
    Min      1Q  Median      3Q     Max 
-35.03 -10.95  -3.45  18.13  31.97 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)   2.7215    26.5522   0.102   0.9196    
set$obstruction 1.1663     0.5355   2.178   0.0447 *  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 20.19 on 16 degrees of freedom
Multiple R-squared:  0.2287,    Adjusted R-squared:  0.1804 
F-statistic: 4.743 on 1 and 16 DF,  p-value: 0.04473

# Tests the normality of model residuals and plots their histogram
shapiro.test(reg_abund$residuals)

```

```

Shapiro-Wilk normality test

data: reg_abund$residuals
W = 0.95713, p-value = 0.5473

```

5.2.3 Plotting the graph

```

plot_abund2 <-
  ggplot(set,
         aes(x = obstruction, y = abundance,
             fill = forest.type)) +
  scale_y_continuous(limits = c(NA, 100),
                     breaks = c(20, 40, 60, 80, 100)) +
  geom_smooth(method = "lm", formula = y ~ x,
              se = TRUE, color = "black",

```

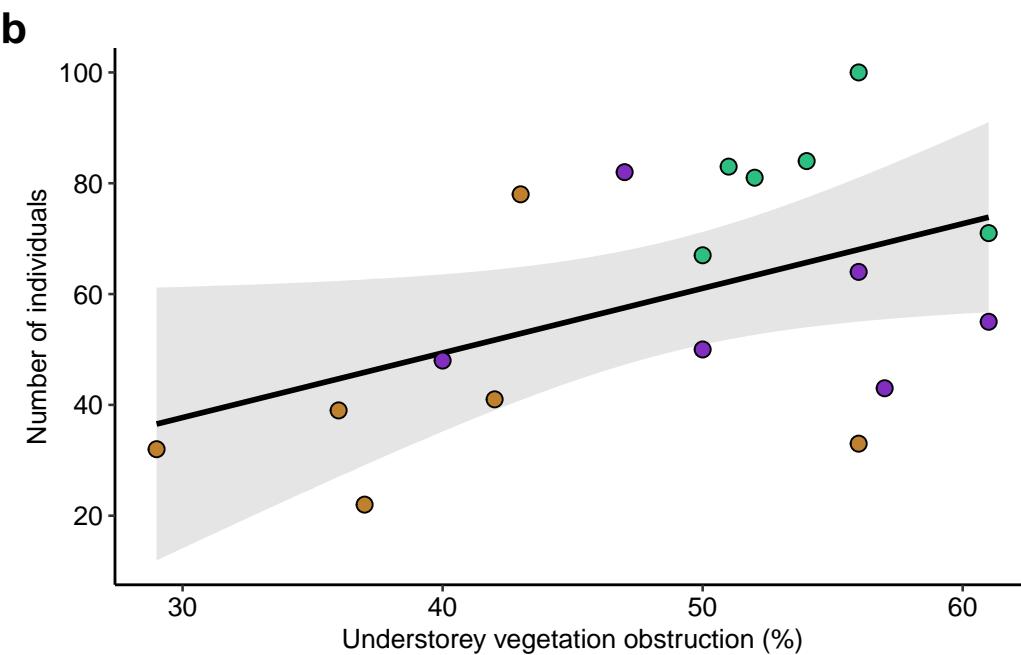
```

    fill = "gray", fullrange = TRUE) +
  geom_point(colour = "black", size = 2.5, shape = 21) +
  scale_fill_manual(values = c("#812DBF", "#2DBF81", "#BF812D"),
                    labels = c("Araucaria plantation", "Natural forest",
                              "Pine plantation")) +
  labs(x = "Understorey vegetation obstruction (%)",
       y = "Number of individuals",
       tag = "b") +
  theme_pubr(base_size = 10) +
  theme(legend.position = "none") +
  theme(plot.tag = element_text(size = 16, face = "bold"))

```

5.2.4 Saving the graph

```
plot_abund2
```



```

ggsave(
  plot = plot_abund2,
  filename = "abundance_understory.png", dpi = 500,
  width = 7, height = 7, units = 'cm')

```

5.3 Indicator species

5.3.1 Preparing data

```

# Selects only indicator species
abund <- abund[,c(1,6,21,38,48,53)]

forest <- abund[1:6, -1]
forest <- colSums(forest)
forest <- as.data.frame(forest)
colnames(forest)[1] <- "forest.type"

araucaria <- abund[7:12, -1]
araucaria <- colSums(araucaria)
araucaria <- as.data.frame(araucaria)
colnames(araucaria)[1] <- "forest.type"

pinus <- abund[13:18, -1]
pinus <- colSums(pinus)
pinus <- as.data.frame(pinus)
colnames(pinus)[1] <- "forest.type"

# Creates new data.table
forest.type <- rbind(forest, araucaria, pinus)
colnames(forest.type)[1] <- "abundance"

forest.type$forest.type <- NA
forest.type$forest.type[1:5] <- "3_natural_forest"
forest.type$forest.type[6:10] <- "2_araucaria_plantation"
forest.type$forest.type[11:15] <- "1_pine_plantation"

forest.type$species <- rownames(forest.type)
forest.type <- forest.type[,c(3,1,2)]

```

```

rownames(forest.type) <- NULL

forest.type$species[c(1,6,11)] <- "4_Chiroxiphia_caudata"
forest.type$species[c(2,7,12)] <- "5_Leptopogon_amaurocephalus"
forest.type$species[c(3,8,13)] <- "3_Synallaxis_cinerascens"
forest.type$species[c(4,9,14)] <- "1_Turdus_leucomelas"
forest.type$species[c(5,10,15)] <- "2_Zonotrichia_capensis"

```

5.3.2 Plotting the graph

```

plot_indsp <-
  ggplot(forest.type,
         aes(x = species, y = abundance, fill = forest.type)) +
  
  scale_fill_manual(values = c("3_natural_forest" = "#2DBF81",
                               "2_arauaria_plantation" = "#812DBF",
                               "1_pine_plantation" = "#BF812D"),
                     labels = c("1_pine_plantation" = "Pine plantation",
                               "2_arauaria_plantation" = "Araucaria plantation",
                               "3_natural_forest" = "Natural forest")) +
  
  geom_bar(stat = "identity", position = "fill",
            alpha = 0.30,
            colour = "black",
            width = 0.7) +
  
  coord_flip() +
  
  scale_y_continuous(labels = scales::percent) +
  
  scale_x_discrete(labels = c("5_Leptopogon_amaurocephalus" = "Leptopogon amaurocephalus",
                             "4_Chiroxiphia_caudata" = "Chiroxiphia caudata",
                             "3_Synallaxis_cinerascens" = "Synallaxis cinerascens",
                             "2_Zonotrichia_capensis" = "Zonotrichia capensis",
                             "1_Turdus_leucomelas" = "Turdus leucomelas")) +
  
  labs(x = " ", y = "Individuals captured by forest type", fill = " ") +
  
  theme_pubr(base_size = 12) +

```

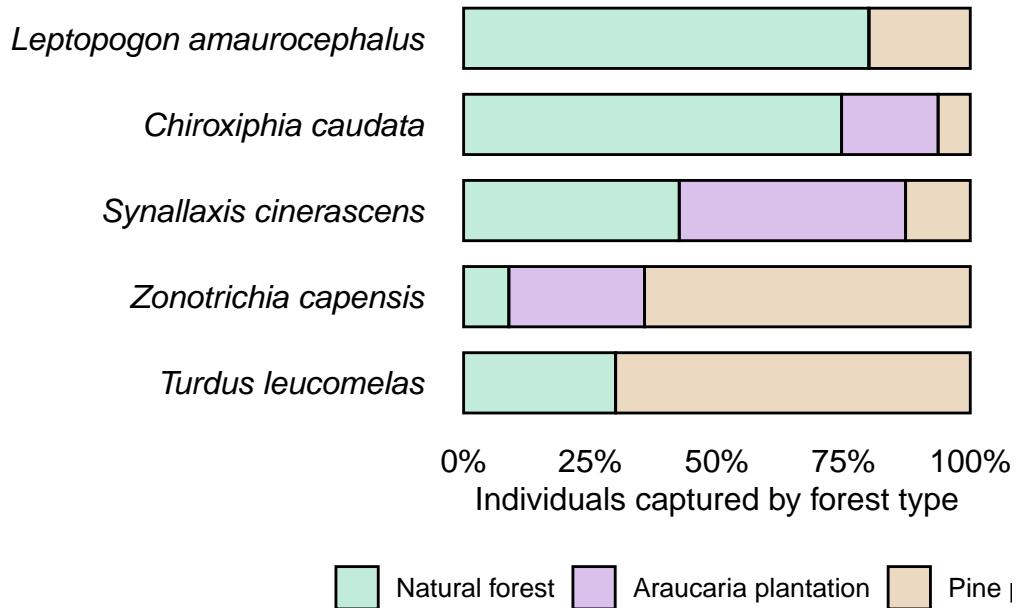
```

theme(panel.border = element_blank(),
      axis.line = element_blank(),
      axis.ticks = element_blank(),
      legend.position = "bottom",
      axis.text.y = element_text(face = "italic")) +
guides(fill = guide_legend(reverse = TRUE))

```

5.3.3 Saving the graph

```
plot_indsp
```



```

ggsave(
  plot = plot_indsp,
  filename = "abund_sp_indicadoras.png", dpi = 500,
  width = 7*3, height = 7, units = 'cm')

```

6 Bird Community Integrity

Predictor variables:

1. **Forest type** - We performed an ANOVA test followed by a Tukey's post hoc test to identify which forest types exhibited significant differences in community integrity from each other.
2. **Understorey vegetation obstruction** - We performed simple linear regression analyses to investigate the effect of understorey vegetation obstruction on community integrity.

Ensuring Reproducibility

```
set.seed(26)
```

Load packages

```
library(ggplot2)
library(ggpubr)
library(vegan)
library(dplyr)
```

Importing data

```
# https://knb.ecoinformatics.org/view/doi:10.5063/F1QR4VK1
setwd("C:/Users/ivana/OneDrive/FLONA_aves/data")
data <- read.csv("FLONA-PF_dataset.txt", sep = "\t")
```

Preparing data

```

# Renaming forest types for consistent ordering in the plot
data$forest.type <- recode(data$forest.type,
                           "Natural_forest" = "1.Natural_forest",
                           "Araucaria_plantation" = "2.Araucaria_plantation",
                           "Pine_plantation" = "3.Pine_plantation")

abund <- data[,c(9:59)]
rel.abund <- abund/rowSums(abund)

```

6.1 Forest type

6.1.1 Calculating Community Integrity

```

# Defines the row indices for control sites representing natural forest
control.rows <- c(1:6)

# Calculates the Bray-Curtis distance matrix and extracts the values among control sites (natural)
bray.cf <- as.matrix(vegdist(rel.abund, method = "bray"))[control.rows, control.rows]

# Calculates the community integrity for control sites as the average similarity (1 minus mean)
int.cf <- mean(1 - colMeans(bray.cf))

# Calculate the Bray-Curtis index between all sites (control and treatment)
bray.all <- as.matrix(vegdist(rel.abund, method = "bray"))[control.rows, ]

# Calculate the community integrity index for all sites
# Scaled by the average similarity of the control sites and expressed as a percentage
int.all <- ((1 - colMeans(bray.all)) / int.cf) * 100

# Display the community integrity index for all sites
int.all

```

	1	2	3	4	5	6	7	8
101.61364	94.01150	103.93656	103.00584	98.10394	99.32853	91.83015	76.08282	
9	10	11	12	13	14	15	16	
84.77477	75.14153	98.01282	89.30392	75.94806	53.72609	55.53405	54.43151	
17	18							
79.95784	57.49437							

6.1.2 Calculating ANOVA

```
# Performs a one-way ANOVA to test the effect of forest type on community integrity and check  
anova <- aov(int.all ~ data$forest.type)  
summary(anova)
```


	Df	Sum Sq	Mean Sq	F value	Pr(>F)						
data\$forest.type	2	4219	2109.6	26.94	1.08e-05 ***						
Residuals	15	1174	78.3								

Signif. codes:	0	'***'	0.001	'**'	0.01	'*'	0.05	'. '	0.1	' '	1

```
shapiro.test(anova$residuals)
```

```
Shapiro-Wilk normality test  
  
data: anova$residuals  
W = 0.93721, p-value = 0.2594
```

The residuals of the ANOVA model are normally distributed ($p > 0.05$), so it is appropriate to proceed with this test to compare species composition between forest types.

```
# Performs Tukey's HSD test to compare pairwise differences between forest types  
TukeyHSD(anova)
```

```
Tukey multiple comparisons of means  
95% family-wise confidence level  
  
Fit: aov(formula = int.all ~ data$forest.type)  
  
$`data$forest.type`  
            diff      lwr      upr  
2.Araucaria_plantation-1.Natural_forest -14.14233 -27.41207 -0.8725946  
3.Pine_plantation-1.Natural_forest     -37.15135 -50.42109 -23.8816100  
3.Pine_plantation-2.Araucaria_plantation -23.00902 -36.27875 -9.7392777  
                                         p adj  
2.Araucaria_plantation-1.Natural_forest 0.0360896  
3.Pine_plantation-1.Natural_forest     0.0000077  
3.Pine_plantation-2.Araucaria_plantation 0.0011497
```

The results showed that the bird communities in the understorey differed significantly between the three forest types (ANOVA: $F(2,15) = 26.94$, $p < 0.001$). Pairwise comparisons revealed that both plantation types had different bird communities compared to the natural forest. There was also a significant difference in community integrity between the two plantation types.

6.1.3 Plotting the graph

```
tab <- data.frame(forest.type = data$forest.type,
                   integrity = int.all)

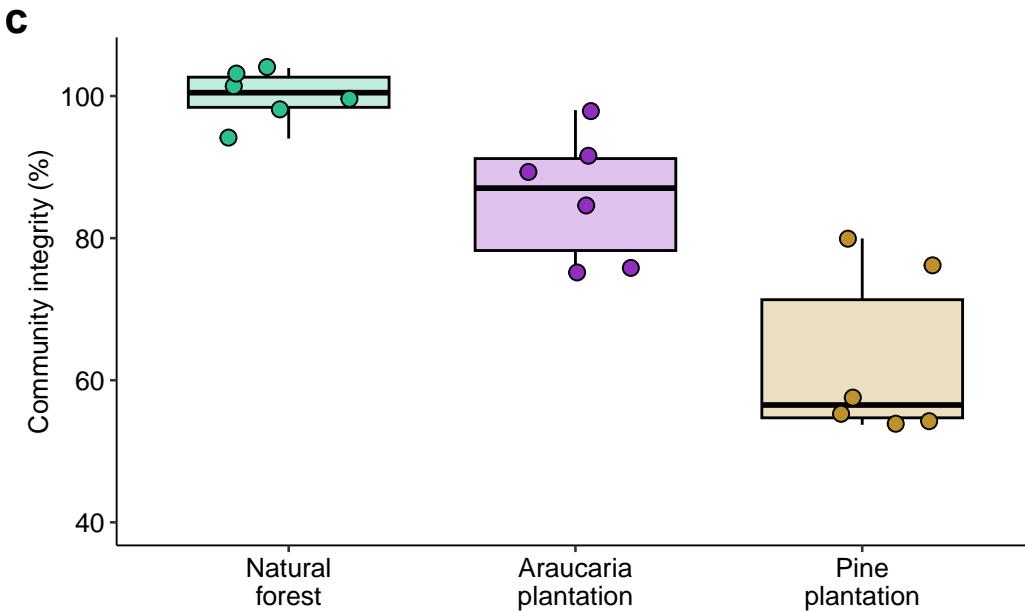
# Creates a vector of colours, assigning one colour to each forest type (6 samples per type)
color <- rep(c("#2DBF91", "#912DBF", "#BF912D"), each = 6)

plot_comp1 <-
  ggplot(data = tab,
         aes(x = forest.type, y = integrity)) +
  geom_boxplot(outlier.shape = NA,
               width = 0.7, col = "black",
               fill = c("#2DBF91", "#912DBF", "#BF912D"),
               alpha = 0.3, show.legend = FALSE) +
  geom_jitter(shape= 21,
              width = 0.25, height = 0.3, colour = "black",
              fill = color, size = 2.5) +
  scale_y_continuous(limits = c(40, 105),
                     breaks = c(40,60,80,100)) +
  scale_x_discrete(labels = c("Natural\nforest", "Araucaria\nplantation",
                             "Pine\nplantation")) +
  labs(x = " ", y = "Community integrity (%)", colour="black",
       tag = "c") +
  theme_pubr(base_size = 10) +
  theme(axis.line = element_line(linewidth = 1/3)) +
```

```
theme(plot.tag = element_text(size = 16, face = "bold"))
```

6.1.4 Saving the graph

```
plot_compl
```



```
ggsave(  
  plot = plot_compl,  
  filename = "composition_foresttype.png", dpi = 500,  
  width = 7, height = 7, units = 'cm')
```

6.2 Understory vegetation obstruction

6.2.1 Preparing data

```
# Creates a new data frame with forest type, species composition, and understorey vegetation
set <- data.frame(forest.type = data$forest.type,
                  int.all = int.all,
                  obstruction = data$obstruction)
```

6.2.2 Fitting the regression model

```
# Fits a linear model to assess the effect of vegetation obstruction on species composition
reg_int = lm(set$int.all ~ set$obstruction)
```

```
# Displays a summary of the model
summary(reg_int)
```

Call:

```
lm(formula = set$int.all ~ set$obstruction)
```

Residuals:

Min	1Q	Median	3Q	Max
-33.188	-10.740	4.382	12.175	17.026

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	30.3533	20.1200	1.509	0.1509
set\$obstruction	1.0773	0.4058	2.655	0.0173 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 15.3 on 16 degrees of freedom

Multiple R-squared: 0.3058, Adjusted R-squared: 0.2624

F-statistic: 7.048 on 1 and 16 DF, p-value: 0.0173

```
# Tests the normality of model residuals
shapiro.test(reg_int$residuals) # It is normal (p > 0.05)
```

Shapiro-Wilk normality test

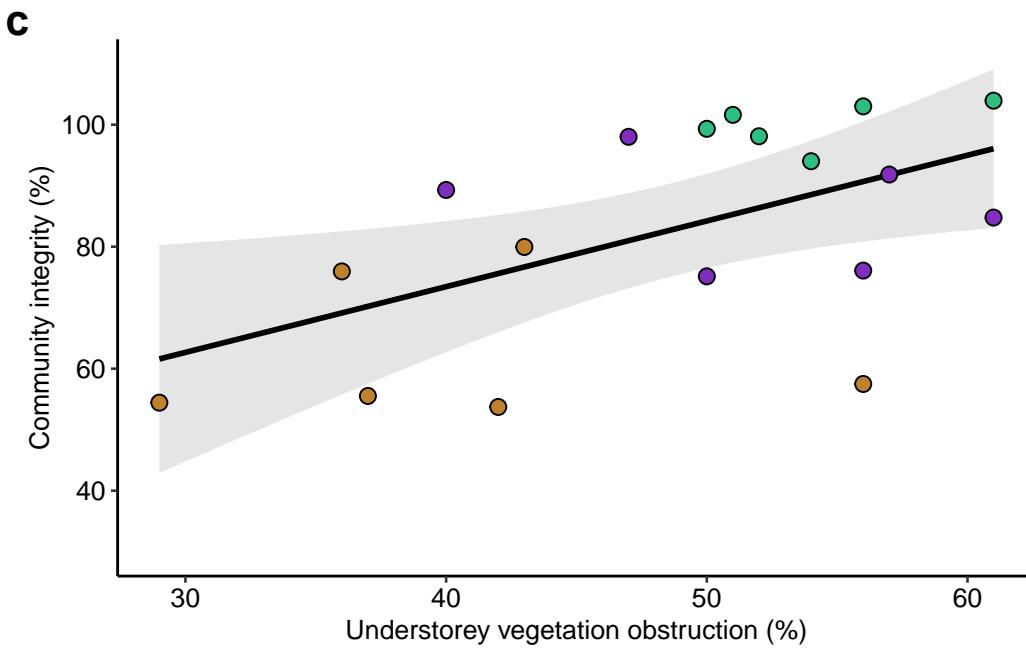
```
data: reg_int$residuals
W = 0.91843, p-value = 0.1213
```

6.2.3 Plotting the graph

```
plot_comp2 <-
  ggplot(set, aes(x = obstruction, y = int.all,
                  fill = forest.type)) +
  scale_y_continuous(limits = c(30, 110),
                     breaks = c(40,60,80,100)) +
  geom_smooth(method = "lm", formula = y ~ x, se = TRUE, color = "black", fill = "gray") +
  geom_point(colour = "black", size = 2.5, shape = 21) +
  scale_fill_manual(values = c("#2DBF81", "#812DBF", "#BF812D"),
                    labels = c("Natural forest", "Araucaria plantation",
                              "Pine plantation")) +
  labs(x = "Understorey vegetation obstruction (%)", y = "Community integrity (%)",
       tag = "c") +
  theme_pubr(base_size = 10) +
  theme(legend.position = "none") +
  theme(plot.tag = element_text(size = 16, face = "bold"))
```

6.2.4 Saving the graph

```
plot_comp2
```



```
ggsave(  
  plot = plot_comp2,  
  filename = "composition_understory.png", dpi = 500,  
  width = 7, height = 7, units = 'cm')
```