[DRAFT] Fire severity impact on understory vegetation species richness in the Araucaria-Nothofagus forest.

Finley, Shersten; Fitzgerald, Avarie; Pham, Peter; Arroyo-Vargas, Paola

Notes:

1. Resource: https://www.elsevier.com/connect/11-steps-to-structuring-a-science-paper-editors-will-takeseriously
2. [To Do] Decrease margins
3. Standardized plot colors: H - "Firebrick", L - "darkbrickedgold3", UN - "darkseagreen"

Abstract

- Brief summary of entire article
- 1 paragraph (5-7 sentences)
- What has been done?
- What are the main findings?

Increased understanding of wildfire impact is critical to understanding the biological systems that determine forest understory succession and diversification.

Post-wildfire plant richness and abundance was collected using 1 x 1 meter quadrat sampling within the Araucaria-Nothofagus forest to produce 220 observations across 8 variables.

Analysis of variance conducted between three burn categories - unburned (UN), low burn (L), and high burn(H) - show that at least one group is significantly different from each other at the 95% confidence level (p-value = $4.359x10^{-15}$).

Post-hoc analysis via Tukey's honestly significant difference further distinguishes that each group was significantly different from each other at the 95% confidence interval (p-value: UN-L = $2.9x10^{-5}$; UN-H = $5.3x10^{-16}$, L-H = $2.2x10^{-6}$).

At the 95% confidence level, best subset linear modeling suggests that [name some variables] have the strongest effects on post-wildfire species diversity.

Wildfire impact is concluded to significantly decrease plant species richness of Araucaria-Nothofagus forest understory communities.

Introduction

- No results
- Last paragraph of the introduction must be the hypothesis/objective.
- Contextual background of the problem and our approach
- State the purpose of the paper and research strategy adopted to answer the question.
- Are there any existing solutions/knowledge? (include references and acknowledgments here)
- Limitations of the question?
- What do we hope to achieve?
- Why is this important?
- Add references

Fire has dramatic impacts on ecosystems and the landscape, retaining observable effects for many years afterward. In ecosystems where fire has not been a frequent driver of ecological change or an evolutionary pressure, as in the case of the Araucaria-Nothofagus of Chile, this process can potentially cause considerable destruction—especially as climate change exacerbates the occurrence and severity of wildfires.

The Araucaria (Araucaria araucana) and Nothofagus (Nothofagus spp) trees that characterize this forest type can survive low-severity wildfires due to branch height adaptations but lack many of the adaptations (serotiny, bark structure??) typified by Pinus-dominated forests in response to fire-driven change. It is likely that the impact of fire may drastically change an entire forest structure, and with that in mind, this analysis attempts to investigate the question of how fire severity affects understory species richness in the Araucaria-Nothofagus forests of Chile. Fire severity's effect on biodiversity serves as a useful tool for forest management.

We look to determine the major factors that influence understory species forests of Araucaria-Nothofagus forest by comparing the variances between plots as well as developing a model to describe the system. Severity is categorized as three groups: unburned, low severity, and high severity. We look to test to the differences in the mean unique species counted in each category to see if there exists a significant difference between each category.

Main null and alternate hypotheses?

Methods

Dataset

- Collection methods
- If we combined datasets, how did we do it? What criteria was it based upon?

Plant richness and abundance was collected using quadrat sampling (1 x 1 m) inside of 55 circular plots (radius = 15 m) within a Araucaria-Nothofagus forest from REGION? of Chile. Each quadrant was 7 m away from the plot center towards each cardinal direction from where species richness and abundance were observed and recorded by a field researcher. The dataset contained 1077 observations across 15 variables that included categorical, continuous, and geospatial variables. For the scope of analysis, the outcome of species richness due to burn severity was of primary interest. Thus, observations were grouped to produce the total number of unique species per quadrant. Geospatial and elevation data were kept for regression model fitting. The resulting dataset contained 220 observations across 8 usable variables.

For the scope of analysis, we are particularly interested in the outcome of species richness due to burn severity. Thus, observations were grouped to produce the total number of unique species per plot. Geospatial and elevation data were kept for modeling fitting. The resulting dataset contains 220 observations across 8 usable variables. Please see the table below for more details.

Visual Description of the data

In figure 1, below, we see comparisons of each burn group's unique species distribution. The boxplots, left, show the differences in means between each group. As burn severity increases in category, unique species count decreases. The plots support the assumption that the frequency of observed species in each plot are approximately normally distributed.

- Hard to differentiate colors in the histogram

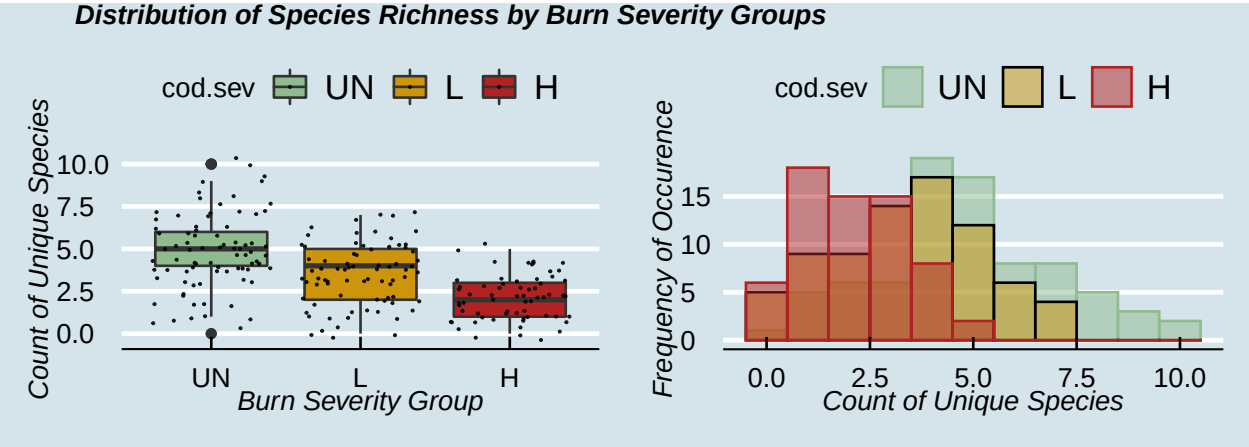


Figure 1: Distributions of Unique Specices by Burn Severity

Histograms, figure 1 right, shows each burn category's distribution relative to each other. Each group is tightly clustered around each other and further support the notion of a normally distribution given the symmetrical shape of each histogram's density curves.

- GIS plots (send to appendix)
- Graphs of sampling method (send to appendix)

Numerical Description of the dataset

- Brief description of dataset that coincides with the visual description. Tie in both the table and plots to a singulgar description of the data.

Numerical description of the data distribution is detailed in table 1. Here, it's clear that each unique species count in category is ranged from 0 to 10. Additionally, each category's mean and median values are not exceptionally different, indicating symmetricality in each burn category's distribution as well as the total.

Table 1: Summary Statistics for Species Richness by Burn Severity

Burn Severity	Minimum	Q1	Median	Mean	S.Error	S.D.	Q3	Maximum
UN	0	4	5	4.84	0.25	2.20	6	10
L	0	2	4	3.43	0.21	1.86	5	7
H	0	1	2	2.11	0.16	1.30	3	5
All	0	2	4	3.56	0.15	2.15	5	10

The absolute differences between the medians and means are 0.16, 1.57, and 0.11 species for unburned, low-burn, and highly burned plots, respectively. Each of these differences are within the standard deviations for each respective group (see table 1; S.D.). The close groupings of these two metrics supports symetrically in species observations and further supports normality. Though unburned plots had higher counts of unique

species, it also had the largest amount of variance out of all burn categories (UN Standard Deviation = 2.20 species).

2.1 Analysis of Variance

This dataset is a prime candidate for analysis of variance (ANOVA) testing. The variable of interest - species richness - is a continuous variable that relates directly to treatment great, in this cas, burn severity. The three treatment groups for burn severity are high, low, and unburned. However, as references in table 1.1, each burn group has an unequal number of observations. This will likely skew the results of our test output. Thus, we proceed with conducting an ANOVA that is followed by using Tukey's HSD for post-hoc analysis.

We propose the sample hypothesis:

H_0: There does not exist a difference in species richness between burn severity groups
H_1: There does exist a difference in species richness between burn severity groups

2.2 Verification of Assumptions

ANOVA testing requires three assumptions in order to be conducted:

1. Observations are independently and randomly selected from a population
2. The data of each group are normally distributed
3. The data have equal and constant variance

2.3 Normality Assumption

- Plots
- Shapiro-Wilk's test
- Add density curves

A primary assumption for analysis of variance testing requires that the dataset is normally distributed. The symmetrical shapes of each burn group's density curve (figure 2) eludes to normality in the count of unique species in each burn category. It should be noted that the high burn group is slightly right skewed.

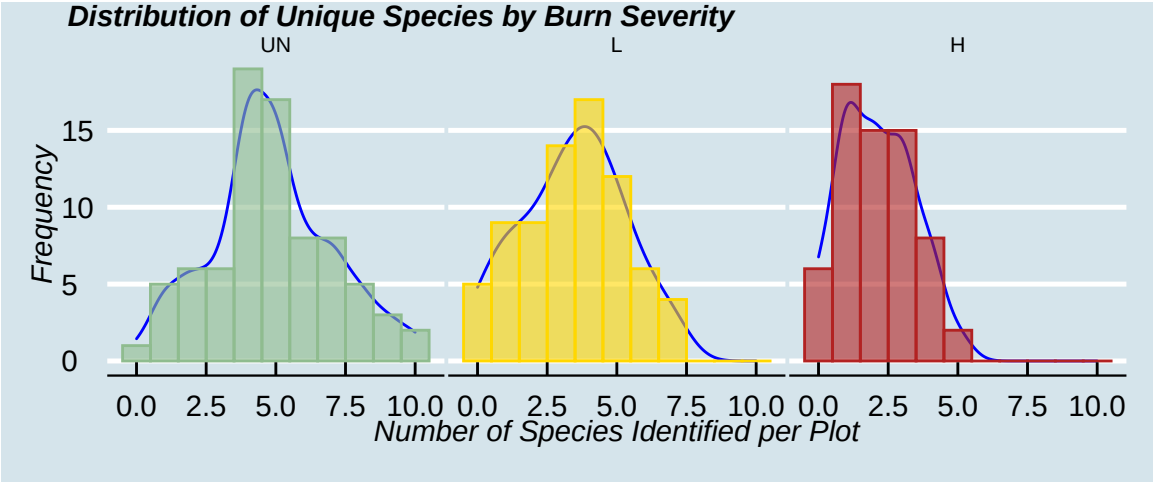


Figure 2: Indidividual Distribution of Each Burn Category

We numericall confirm each burn group's normality by using Shapiro-Wilk's test. We define the null hypothesis for the Shapiro-Wilks test:

H_O = The data is normally distributed

Table 2: Shapiro-Wilks Test for All Groups

Burn Severity	N Observations	P-Value
UN	80	0.97
L	76	0.96
H	64	0.93

And, conduct the results below, (table 2: Shapiro-Wilks Tests for All groups):

We receive p-values of 0.97, 0.96, and 0.93 for unburned, low-burned, and high-burned groups, respectively (table 2, above). At the 95% confidence level ($\alpha = 0.05$), each p-value is greater than 0.05 and we fail to reject the null hypothesis for each group. Thus, we conclude that each burn group's species count is normally distributed and we have satisfied the first condition for analysis of variance testing.

2.2 Equal Variance Assumption

- Plots
- Levene's Test

An additional assumption for analysis of variance testing requires that the dataset is to contain equal variance from point to point. Brief visual analysis shows the moving leverage points for each plot.

Conducting Levene's F-Test, we confirm if each group has equal variance through the dataset.

Model Selection

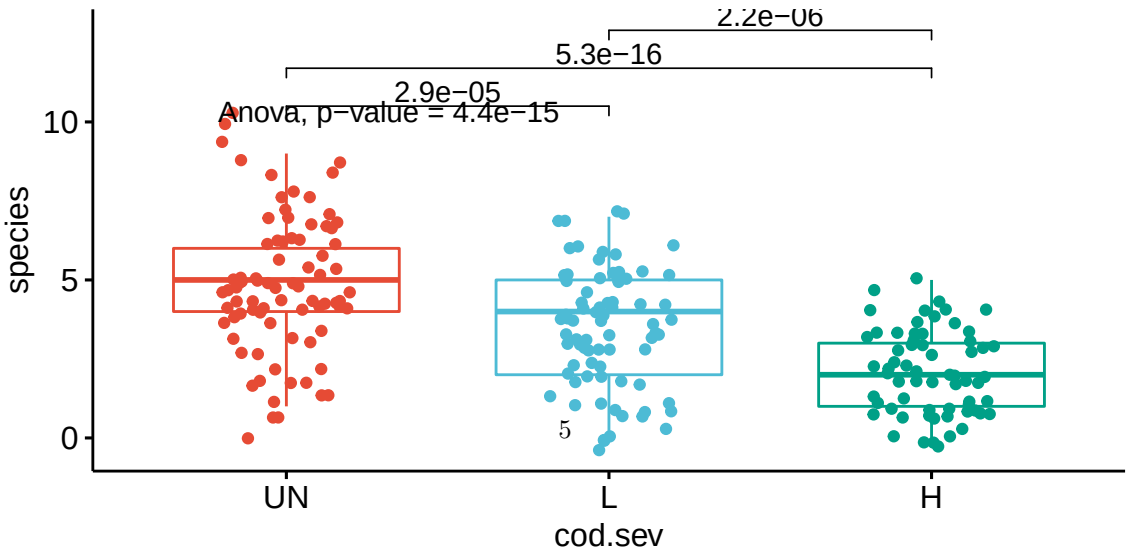
ANOVA testing lends itself well into multiple linear regression. In fact, they are the same procedure. Though we are primarily interested in the differences between burn groups, developing a descriptive model would expand our understanding of the factors at play in this forest system. We proceed to fit and select a best subsets model that most effectively conveys the relationships present in our dataset.

Results

ANOVA & Tukey

Table 3: Analysis of Variance Model

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
cod.sev	2	266.4	133.2	38.66	4.359e-15
Residuals	217	747.8	3.446	NA	NA



Regression Modeling

Conclusion

Discussion

References

Appendix A: Tables & Figures

Appendix B: R Code

```
# Packages Used
library(formatR)
library(broom)
library(plyr)
library(ggpubr)
library(ggthemes)
library(hrbrthemes)
library(kableExtra)
library(knitr)
library(magrittr)
library(tidyverse)
library(xtable)
library(pander)

# Import data from .csv
dt = data.frame(read_csv("ChinaMuerta.csv"))

# Subset "dt" and remove categories that are not "native" or "exotic"
spp <-subset(dt, origen=="native" | origen=="exotic")
spp<-droplevels(spp)

# Quick calculation of the total species richness
# dim(tapply(spp$origen, spp$species, length))  #66 spp

# Reorganize "spp" by species richness by treatment
ddply(spp, c("site","cod.sev"), function(df)
  return(c(species=length(unique(df$species)))))

# Reorganize "spp" further by variables of interest.
plot.rich<-ddply(spp, c("site","n.plot" ,"cod.sev", "transect"), function(df)
  return(c(species=length(unique(df$species)))))

# spread function
# avoiding to lose subplot without species, and filling them with zeros
dt1<-spread(plot.rich, transect, species, fill=0)

# gather function & reorganizing subplots (transects) in one column
dt2<-gather(dt1, "east", "north", "south", "west", key="transect", value=species)

# Set cod.sev as a factor and assign a specific order
dt2$cod.sev = factor(dt2$cod.sev, levels = c("UN","L","H"), ordered=TRUE)

# ---------Create data frame for multiple regression---------
df = dt %>%
  filter(origen != "NA") %>%
  select(n.plot, cod.sev, slope, aspect, elev.m,
         cov.50N, cov.50E, cov.50S, cov.50W, cov.140N, cov.140E, cov.140S,
         cov.140W) %>%
  rowwise() %>%
```

```
  mutate(cov.50mean = mean(c(cov.50N, cov.50E, cov.50S, cov.50W), na.rm=TRUE)) %>%
  mutate(cov.140mean = mean(c(cov.140N, cov.140E, cov.140S, cov.140W), na.rm=TRUE)) %>%
  summarize(n.plot, cod.sev, aspect, slope, elev.m, cov.50mean, cov.140mean)




# Creating a boxplot for the groups
bp = ggplot(dt2, aes(y=species, fill=cod.sev, x=cod.sev)) +
  geom_boxplot() +
  geom_jitter(color="black", size=0.1, alpha=0.9) +
  theme_economist() +
  scale_fill_manual(values=c("darkseagreen","darkgoldenrod3","firebrick")) +
  labs(y = expression(italic("Count of Unique Species")),
       x = expression(italic("Burn Severity Group")),
       subtitle = expression(
         bold(
           "")))

# Creating a histogram for the frequency of the species count
# make the line thicker and fill
hg = ggplot(dt2, aes(x=species, fill = cod.sev, color=cod.sev)) +
  geom_histogram(binwidth=1, alpha=0.5, position="identity")+
  theme_economist() +
  scale_fill_manual(values=c("darkseagreen","darkgoldenrod3","firebrick")) +
  scale_color_manual(values=c("darkseagreen","black","firebrick")) +
  labs(y = expression(italic("Frequency of Occurence")),
       x = expression(italic("Count of Unique Species")),
       subtitle = expression(
         bold(
           "")
       )) +
  theme(legend.position = "top")

# Using ggarrange to combine both plots into a single output
bph = ggarrange(bp, hg, nrow=1)

#A
annotate_figure(bph,
                fig.lab = "Distribution of Species Richness by Burn Severity Groups",
                fig.lab.face="bold.italic",
                fig.lab.pos = "top.left")
# Table 1: Summary Statistics for groups

summary = bind_rows( # We are creating two tables, this function will bind them together

  dt2 %>% # Produce summary statistics by burn group
  dplyr::group_by(cod.sev) %>%
  dplyr::summarise(
    Minimum = min(species),
    Q1 = quantile(species, 0.25),
    Median = median(species),
```

```
    Mean = mean(species),
    S.Error = sd(species)/sqrt(length(species)),
    S.D. = sd(species),
    Q3 = quantile(species, 0.75),
    Maximum = max(species)
    ),

  dt2 %>% # Produce summary statistics for all groups combined
    summarise(
      Minimum = min(species),
      Q1 = quantile(species, 0.25),
      Median = median(species),
      Mean = mean(species),
      S.Error = sd(species)/sqrt(length(species)),
      S.D. = sd(species),
      Q3 = quantile(species, 0.75),
      Maximum = max(species)
      ) %>%
    mutate(cod.sev = "All")
)

# Using KableExtra package to clean up table data and to add styling options
kable(summary,
      caption =
        "Summary Statistics for Species Richness by Burn Severity",
      col.names = c("Burn Severity", names(summary)[-1]),
      align="c",
      booktabs = T,
      digits=2) %>%
  kable_styling(position = "center", latex_options="hold_position") %>%
  row_spec(0, bold=TRUE)

hg.all = dt2 %>%
  mutate(cod.sev = fct_reorder(cod.sev,species, .desc=TRUE)) %>%
  ggplot(aes(x=species, fill=cod.sev, color=cod.sev)) +
  geom_density(color="blue", alpha=0,aes(y=..count..)) +
  geom_histogram(binwidth=1, alpha=0.6, position="identity") +
  scale_fill_manual(values=c("darkseagreen","gold","firebrick")) +
  scale_color_manual(values=c("darkseagreen","gold","firebrick")) +
  theme_economist() +
  theme(
    legend.position="none",
    panel.spacing = unit(0.1, "lines"),
    strip.text.x = element_text(size = 8)
  ) +
  xlab(expression(italic("Number of Species Identified per Plot"))) +
  ylab(expression(italic("Frequency"))) +
  facet_wrap(~cod.sev)

# Annotate Figure
annotate_figure(hg.all,
                fig.lab = "Distribution of Unique Species by Burn Severity",
                fig.lab.face="bold.italic",
```

```
                fig.lab.pos = "top.left")

sw.all = dt2 %>%
  dplyr::group_by(cod.sev) %>%
  dplyr::summarize(
    N = length(species),
    P.Value = shapiro.test(species)$statistic)

# Using KableExtra package to clean up table data and to add styling options
kable(sw.all,
      caption =
        "Shapiro-Wilks Test for All Groups",
      col.names = c("Burn Severity", "N Observations", "P-Value"),
      align="c",
      booktabs = T,
      digits=2) %>%
  kable_styling(position = "center", latex_options="hold_position") %>%
  row_spec(0, bold=TRUE)

# Anova
anova = aov(species~cod.sev, data = dt2)

pander(anova)


# Define variables for a loop
x <- which(names(dt2) == "cod.sev") # name of grouping variable
y <- which(names(dt2) == "species") # names of variables to test
method1 <- "anova" # one of "anova" or "kruskal.test"
method2 <- "t.test" # one of "wilcox.test" or "t.test"
my_comparisons <- list(c("UN", "L"), c("UN", "H"), c("L", "H"))

# Loop function to create and label a boxplot with ANOVA
for (i in y) {
  for (j in x) {
    p <- ggboxplot(dt2,
                   x = colnames(dt2[j]), y = colnames(dt2[i]),
                   color = colnames(dt2[j]),
                   legend = "none",
                   palette = "npg",
                   add = "jitter"
    )
    print(
      p + stat_compare_means(aes(label = paste0(..method.., ", p-value = ", ..p.format..)),
                             method = method1, label.y = max(dt2[, i], na.rm = TRUE)
      )
      + stat_compare_means(comparisons = my_comparisons, method = method2, label = "p.format")
    )
  }
}

# make colors match all of the publication's graphs
```