

# QB Group Project

Ashish Nambiar and Jaeyoung Yoo; Z620: Quantitative Biodiversity, Indiana University

06 March, 2025

## 1. Site-by-species matrix

```
df <- read.csv("lambir_LDW.csv", stringsAsFactors = FALSE)

df$Site_ID <- paste(df$Forest, df$myc_type, df$plot_id, df$canopy_layer, sep = "-")

# Create presence-absence or abundance table
final_df <- df %>%
  group_by(Site_ID, myc_type, Forest, canopy_layer, Latin) %>%
  summarise(count = n(), .groups = "drop") %>%
  pivot_wider(names_from = Latin, values_from = count, values_fill = 0) %>%
  ungroup()

# Set Site_ID as row names and remove metadata columnLatin# Set Site_ID as row names and remove metadata
final_df <- final_df[, -1]
site_species_matrix <- final_df[, -c(1:3)] # Keep only species columns
variable <- final_df[, c(1:3)]
```

## 2. Alpha-Diversity

```
# Calculate richness
richness <- specnumber(site_species_matrix)

# Calculate Simpson's diversity (1-D)
simpson <- simpson.unb(site_species_matrix, inverse = FALSE)

## Warning in rarefy(x, 2): requested 'sample' was larger than smallest site
## maximum (1)

# Calculate Shannon's diversity
#diversity(site_species_matrix, index = "shannon")

#Add richness, Simpson and Shannon in the dataset
variable$Site_ID <- paste(rownames(variable))
rich <- data.frame(Site_ID = rownames(site_species_matrix), Richness = richness)
variable <- variable %>%
  left_join(rich, by = "Site_ID")

simp <- data.frame(Site_ID = rownames(site_species_matrix), Simpson = simpson)
variable <- variable %>%
  left_join(simp, by = "Site_ID")
```

```

#shan <- data.frame(Site_ID = rownames(site_species_matrix), Shannon = shannon)
#variable <- variable %>%
# left_join(shan, by = "Site_ID")

## Plot richness
# Function to summarize data
data_summary <- function(data) {
  data %>%
    group_by(canopy_layer, myc_type) %>%
    summarise(
      mean_richness = mean(Richness, na.rm = TRUE),
      se_richness = sd(Richness, na.rm = TRUE) / sqrt(n()),
      .groups = 'drop'
    )
}

# Function to add "Total" category
data_total <- function(data) {
  data %>%
    group_by(canopy_layer) %>%
    summarise(
      myc_type = "Total",
      mean_richness = mean(Richness, na.rm = TRUE),
      se_richness = sd(Richness, na.rm = TRUE) / sqrt(n()),
      .groups = 'drop'
    )
}

# Filter data for Temperate and Tropical forests
temperate_data <- variable %>% filter(Forest == "Temperate")
temperate_data$canopy_layer <- as.factor(temperate_data$canopy_layer)

# Process data for both forest types
temperate_plot_data <- bind_rows(data_summary(temperate_data), data_total(temperate_data))

# Run ANOVA - Total - Temperate
anova_result1 <- aov(Richness ~ canopy_layer, data = temperate_data)
summary(anova_result1)

##
##           Df Sum Sq Mean Sq F value   Pr(>F)
## canopy_layer    3   28.5    9.505   11.97 9.9e-08 ***
## Residuals  1304 1035.8    0.794
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

tukey_result1 <- TukeyHSD(anova_result1)

# Run ANOVA - AM & EcM - Temperate
anova_result2 <- aov(Richness ~ canopy_layer * myc_type, data = temperate_data)
summary(anova_result2)

##
##           Df Sum Sq Mean Sq F value   Pr(>F)
## canopy_layer    3   28.5    9.505  13.039 2.16e-08 ***
## myc_type        1    0.2    0.192    0.264   0.607
## canopy_layer:myc_type  3   88.0   29.338  40.250 < 2e-16 ***

```

```

## Residuals          1300  947.6   0.729
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

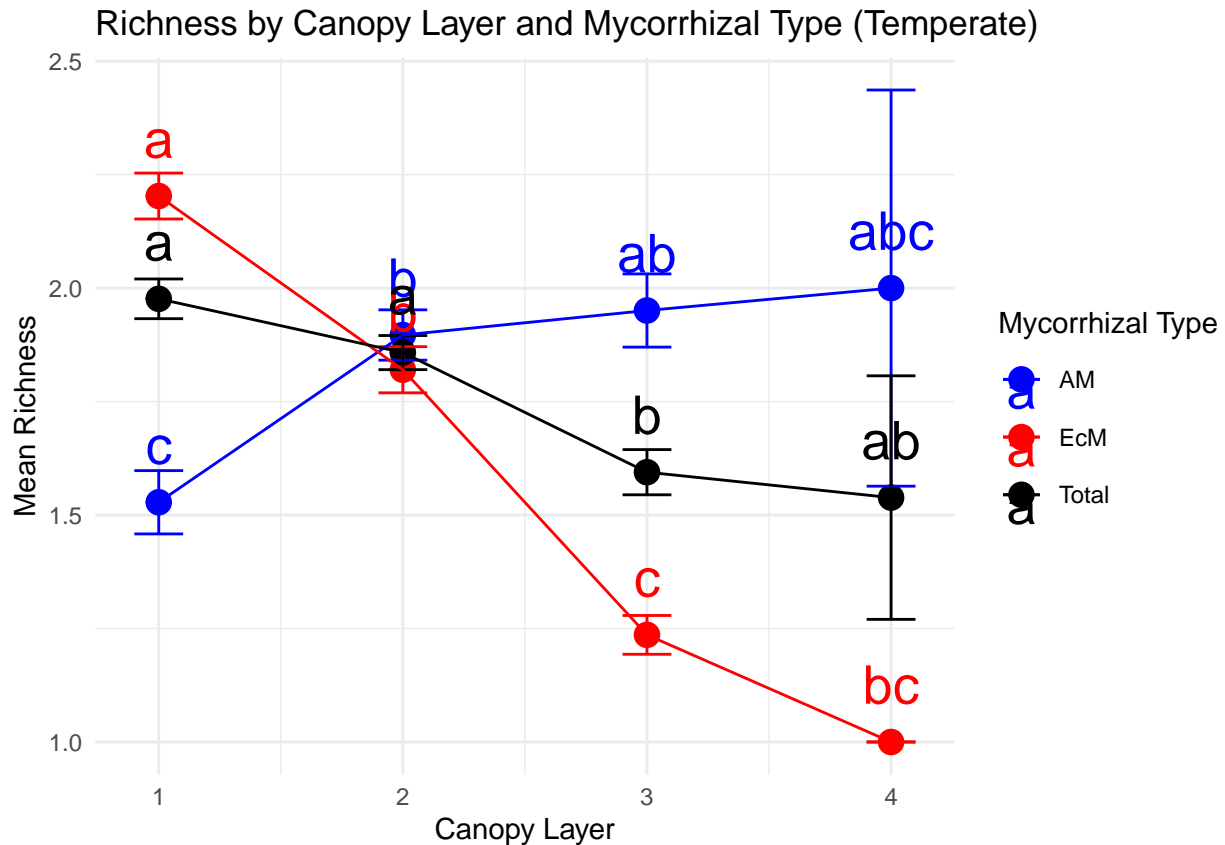
tukey_result2 <- TukeyHSD(anova_result2)

# Create Tukey letters - Temperate
tukey_letters1 <- multcompLetters4(anova_result1, tukey_result1)
tukey_letters2 <- multcompLetters4(anova_result2, tukey_result2)
tukey_df1 <- data.frame(canopy_layer = names(tukey_letters1$`canopy_layer`$Letters),
                        Letters = tukey_letters1$`canopy_layer`$Letters,
                        myc_type = rep("Total"))
tukey_df2 <- data.frame(var = names(tukey_letters2$`canopy_layer:myc_type`$Letters),
                        Letters = tukey_letters2$`canopy_layer:myc_type`$Letters)
tukey_df2$myc_type <- sub(".*:", "", tukey_df2$var)
tukey_df2$canopy_layer <- sub(":.*", "", tukey_df2$var)
tukey_df2 <- tukey_df2[,-1]
tukey_df <- rbind(tukey_df1, tukey_df2)

temperate_plot_data <- merge(temperate_plot_data, tukey_df, by = c("myc_type", "canopy_layer"), all = T)
temperate_plot_data$canopy_layer <- as.numeric(temperate_plot_data$canopy_layer)

# Create plots
ggplot(temperate_plot_data, aes(x = canopy_layer, y = mean_richness, color = myc_type)) +
  geom_point(size = 4) +
  geom_errorbar(aes(ymin = mean_richness - se_richness, ymax = mean_richness + se_richness), width = 0.1) +
  geom_line() +
  # geom_smooth(aes(group = myc_type), method = "lm", se = FALSE, linetype = "solid") +
  theme_minimal() +
  labs(
    title = "Richness by Canopy Layer and Mycorrhizal Type (Temperate)",
    x = "Canopy Layer",
    y = "Mean Richness",
    color = "Mycorrhizal Type"
  ) +
  scale_color_manual(values = c("blue", "red", "black")) +
  geom_text(aes(label = Letters), vjust = -1, size = 7)

```



```
# Repeat for tropical
tropical_data <- variable %>% filter(Forest == "Tropical")
tropical_data$canopy_layer <- as.factor(tropical_data$canopy_layer)

tropical_plot_data <- bind_rows(data_summary(tropical_data), data_total(tropical_data))

# Run ANOVA - Total - Temperate
anova_result1 <- aov(Richness ~ canopy_layer, data = tropical_data)
summary(anova_result1)
```

```
##               Df Sum Sq Mean Sq F value Pr(>F)
## canopy_layer    6 1662798   277133   271.7 <2e-16 ***
## Residuals      5458 5567927    1020
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

tukey_result1 <- TukeyHSD(anova_result1)
```

```
# Run ANOVA - AM & EcM - Temperate
anova_result2 <- aov(Richness ~ canopy_layer * myc_type, data = tropical_data)
summary(anova_result2)
```

```
##               Df Sum Sq Mean Sq F value Pr(>F)
## canopy_layer    6 1662798   277133   768.4 <2e-16 ***
## myc_type         1 2346475 2346475  6505.6 <2e-16 ***
## canopy_layer:myc_type  6 1255368   209228   580.1 <2e-16 ***
## Residuals      5451 1966084    361
## ---
```

```

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

tukey_result2 <- TukeyHSD(anova_result2)

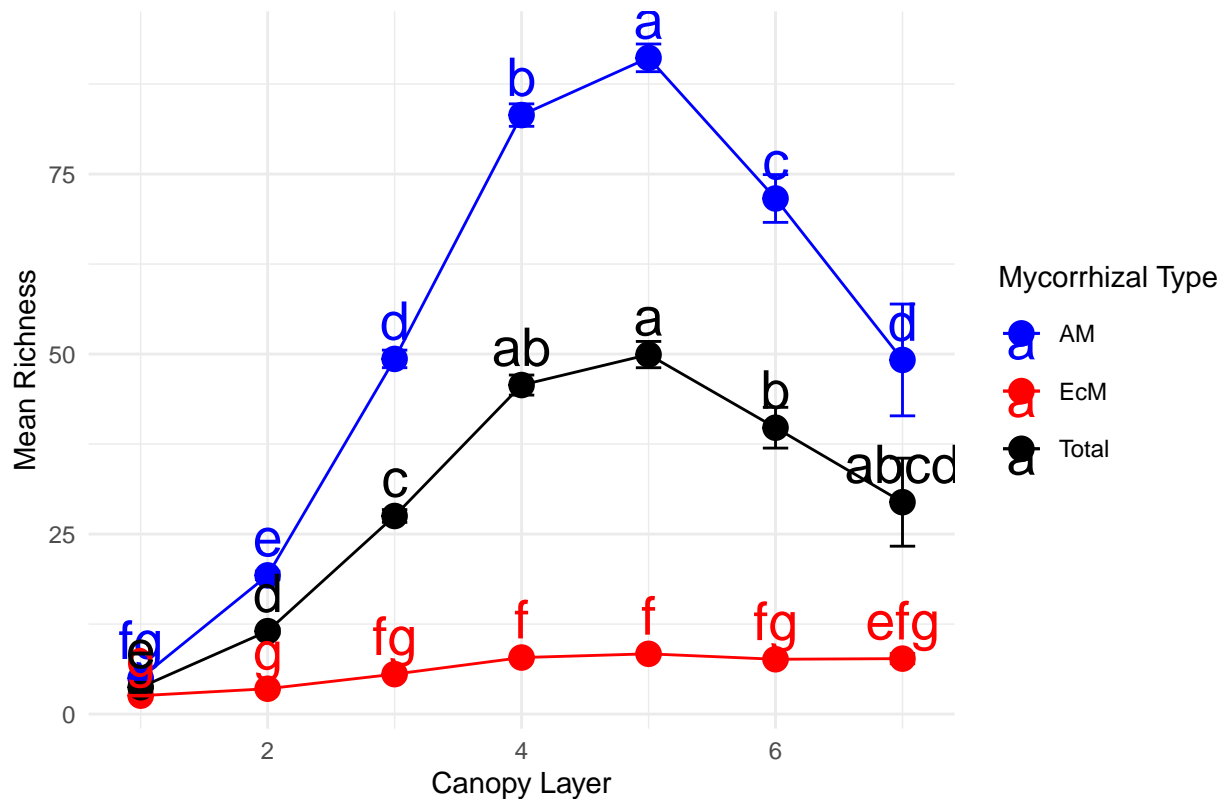
# Create Tukey letters - Temperate
tukey_letters1 <- multcompLetters4(anova_result1, tukey_result1)
tukey_letters2 <- multcompLetters4(anova_result2, tukey_result2)
tukey_df1 <- data.frame(canopy_layer = names(tukey_letters1$`canopy_layer`$Letters),
                       Letters = tukey_letters1$`canopy_layer`$Letters,
                       myc_type = rep("Total"))
tukey_df2 <- data.frame(var = names(tukey_letters2$`canopy_layer:myc_type`$Letters),
                       Letters = tukey_letters2$`canopy_layer:myc_type`$Letters)
tukey_df2$myc_type <- sub(".*:", "", tukey_df2$var)
tukey_df2$canopy_layer <- sub(":.*", "", tukey_df2$var)
tukey_df2 <- tukey_df2[,-1]
tukey_df <- rbind(tukey_df1, tukey_df2)

tropical_plot_data <- merge(tropical_plot_data, tukey_df, by = c("myc_type", "canopy_layer"), all = TRUE)
tropical_plot_data$canopy_layer <- as.numeric(tropical_plot_data$canopy_layer)

# Create plots
ggplot(tropical_plot_data, aes(x = canopy_layer, y = mean_richness, color = myc_type)) +
  geom_point(size = 4) +
  geom_errorbar(aes(ymin = mean_richness - se_richness, ymax = mean_richness + se_richness), width = 0.5) +
  geom_line() +
  # geom_smooth(aes(group = myc_type), method = "lm", se = FALSE, linetype = "solid") +
  theme_minimal() +
  labs(
    title = "Richness by Canopy Layer and Mycorrhizal Type (Tropical)",
    x = "Canopy Layer",
    y = "Mean Richness",
    color = "Mycorrhizal Type"
  ) +
  scale_color_manual(values = c("blue", "red", "black")) +
  geom_text(aes(label = Letters), vjust = -0.5, size = 7)

```

## Richness by Canopy Layer and Mycorrhizal Type (Tropical)



```
## Plot Simpson's diversity
# Function to summarize data
data_summary <- function(data) {
  data %>%
    group_by(canopy_layer, myc_type) %>%
    summarise(
      mean_simpson = mean(Simpson, na.rm = TRUE),
      se_simpson = sd(Simpson, na.rm = TRUE) / sqrt(n()),
      .groups = 'drop'
    )
}

# Function to add "Total" category
data_total <- function(data) {
  data %>%
    group_by(canopy_layer) %>%
    summarise(
      myc_type = "Total",
      mean_simpson = mean(Simpson, na.rm = TRUE),
      se_simpson = sd(Simpson, na.rm = TRUE) / sqrt(n()),
      .groups = 'drop'
    )
}

# Filter data for Temperate and Tropical forests
temperate_data <- variable %>% filter(Forest == "Temperate")
temperate_data$canopy_layer <- as.factor(temperate_data$canopy_layer)
```

```

# Process data for both forest types
temperate_plot_data <- bind_rows(data_summary(temperate_data), data_total(temperate_data))

tropical_data <- variable %>% filter(Forest == "Tropical")
tropical_data$canopy_layer <- as.factor(tropical_data$canopy_layer)

tropical_plot_data <- bind_rows(data_summary(tropical_data), data_total(tropical_data))

# Run ANOVA - Total - Temperate
anova_result1 <- aov(Simpson ~ canopy_layer, data = temperate_data)
summary(anova_result1)

```

```

##              Df Sum Sq Mean Sq F value Pr(>F)
## canopy_layer    3   11.9   3.966   49.44 <2e-16 ***
## Residuals    1304  104.6   0.080
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```

tukey_result1 <- TukeyHSD(anova_result1)

```

```

# Run ANOVA - AM & EcM - Temperate
anova_result2 <- aov(Simpson ~ canopy_layer * myc_type, data = temperate_data)
summary(anova_result2)

```

```

##              Df Sum Sq Mean Sq F value  Pr(>F)
## canopy_layer    3  11.90   3.966  55.567 < 2e-16 ***
## myc_type         1   0.55   0.550   7.711 0.00557 **
## canopy_layer:myc_type  3  11.27   3.757  52.634 < 2e-16 ***
## Residuals      1300  92.78   0.071
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```

tukey_result2 <- TukeyHSD(anova_result2)

```

```

# Create Tukey letters - Temperate
tukey_letters1 <- multcompLetters4(anova_result1, tukey_result1)
tukey_letters2 <- multcompLetters4(anova_result2, tukey_result2)
tukey_df1 <- data.frame(canopy_layer = names(tukey_letters1$`canopy_layer`$Letters),
                      Letters = tukey_letters1$`canopy_layer`$Letters,
                      myc_type = rep("Total"))
tukey_df2 <- data.frame(var = names(tukey_letters2$`canopy_layer:myc_type`$Letters),
                      Letters = tukey_letters2$`canopy_layer:myc_type`$Letters)
tukey_df2$myc_type <- sub(".*:", "", tukey_df2$var)
tukey_df2$canopy_layer <- sub(":.*", "", tukey_df2$var)
tukey_df2 <- tukey_df2[,-1]
tukey_df <- rbind(tukey_df1, tukey_df2)

```

```

temperate_plot_data <- merge(temperate_plot_data, tukey_df, by = c("myc_type", "canopy_layer"), all = T)
temperate_plot_data$canopy_layer <- as.numeric(temperate_plot_data$canopy_layer)

```

```

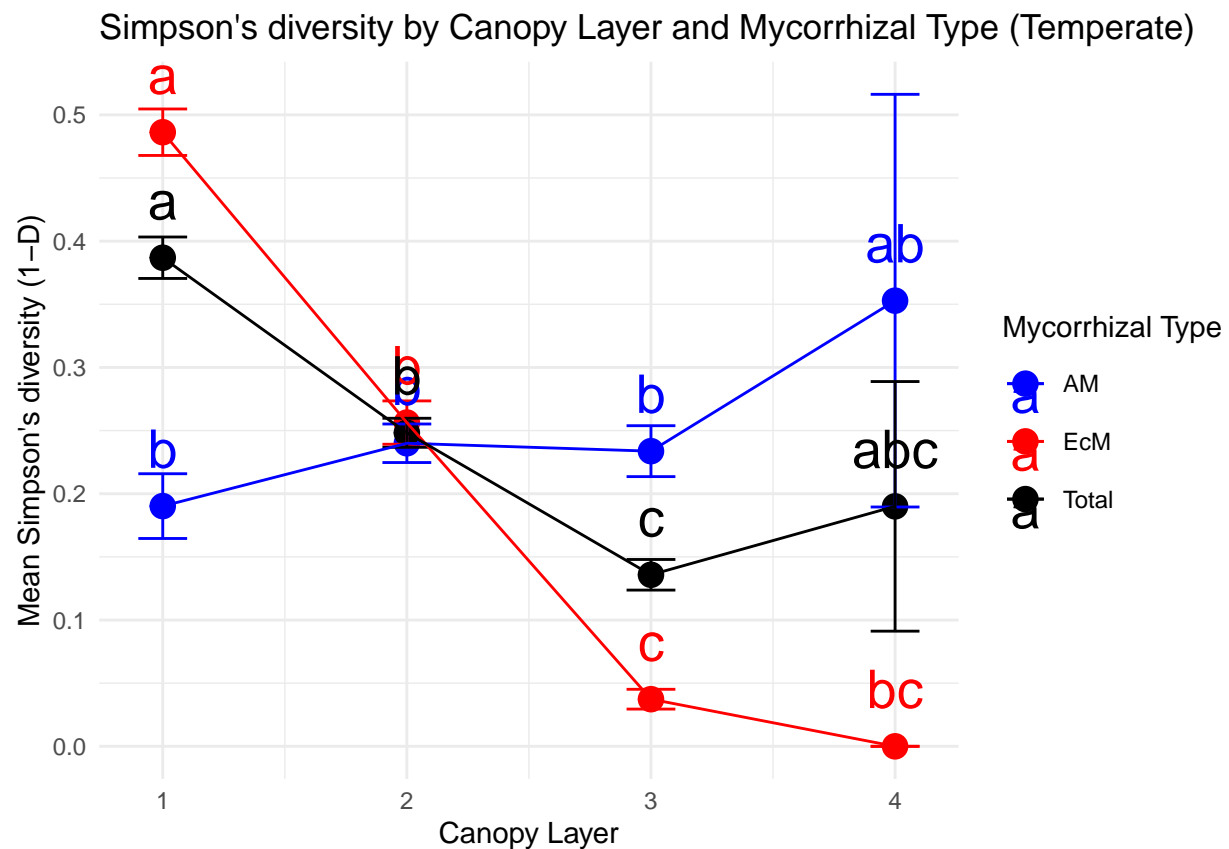
# Create plots
ggplot(temperate_plot_data, aes(x = canopy_layer, y = mean_simpson, color = myc_type)) +
  geom_point(size = 4) +
  geom_errorbar(aes(ymin = mean_simpson - se_simpson, ymax = mean_simpson + se_simpson), width = 0.2) +

```

```

geom_line() +
# geom_smooth(aes(group = myc_type), method = "lm", se = FALSE, linetype = "solid") +
theme_minimal() +
labs(
  title = "Simpson's diversity by Canopy Layer and Mycorrhizal Type (Temperate)",
  x = "Canopy Layer",
  y = "Mean Simpson's diversity (1-D)",
  color = "Mycorrhizal Type"
) +
scale_color_manual(values = c("blue", "red", "black")) +
geom_text(aes(label = Letters), vjust = -1, size = 7)

```



```

# Repeat for tropical
tropical_data <- variable %>% filter(Forest == "Tropical")
tropical_data$canopy_layer <- as.factor(tropical_data$canopy_layer)

tropical_plot_data <- bind_rows(data_summary(tropical_data), data_total(tropical_data))

# Run ANOVA - Total - Tropical
anova_result1 <- aov(Simpson ~ canopy_layer, data = tropical_data)
summary(anova_result1)

```

```

##           Df Sum Sq Mean Sq F value Pr(>F)
## canopy_layer    6   12.6   2.0928   35.12 <2e-16 ***
## Residuals  5458   325.3   0.0596
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```



```

tukey_result1 <- TukeyHSD(anova_result1)

# Run ANOVA - AM & EcM - Tropical
anova_result2 <- aov(Simpson ~ canopy_layer * myc_type, data = tropical_data)
summary(anova_result2)

##               Df Sum Sq Mean Sq F value    Pr(>F)
## canopy_layer      6  12.56    2.09   40.57 < 2e-16 ***
## myc_type          1  39.65   39.65  768.54 < 2e-16 ***
## canopy_layer:myc_type  6   4.44    0.74   14.35 2.65e-16 ***
## Residuals       5451 281.19    0.05
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

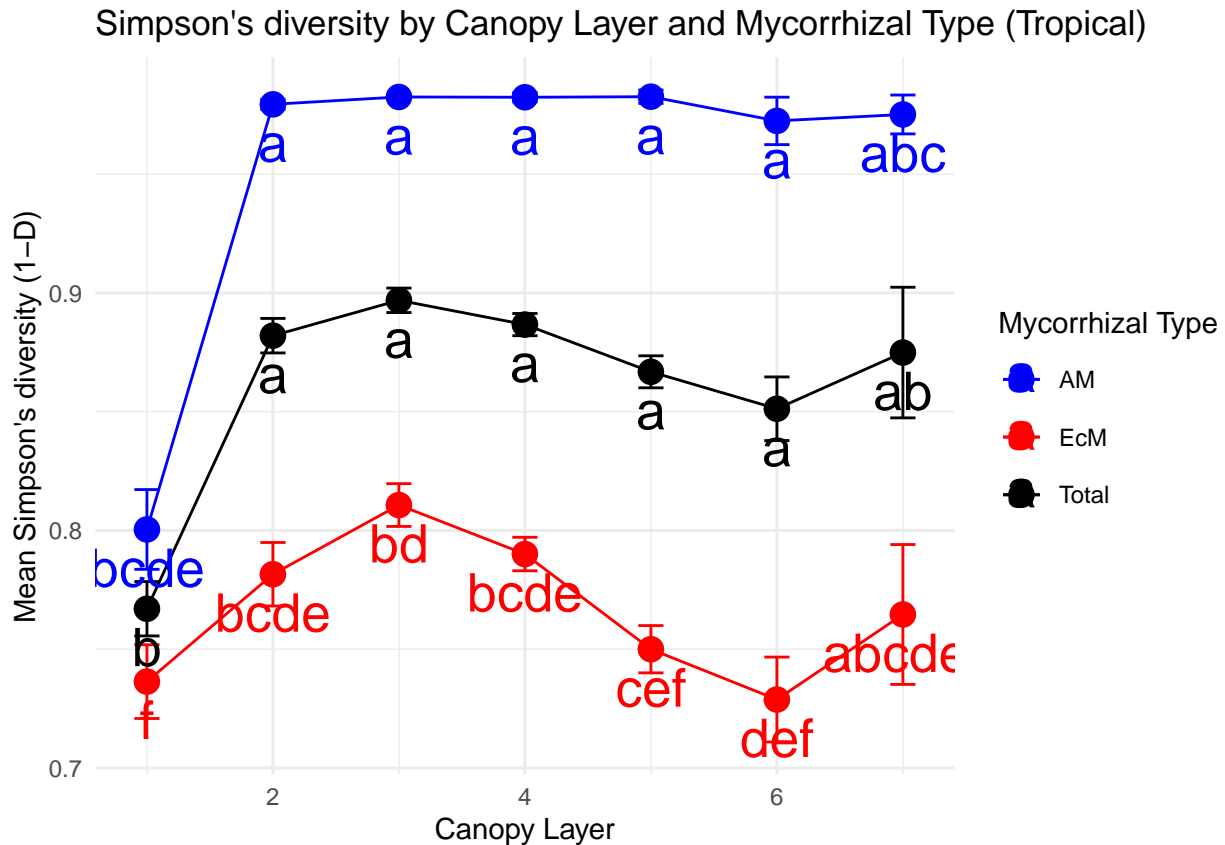
tukey_result2 <- TukeyHSD(anova_result2)

# Create Tukey letters - Tropical
tukey_letters1 <- multcompLetters4(anova_result1, tukey_result1)
tukey_letters2 <- multcompLetters4(anova_result2, tukey_result2)
tukey_df1 <- data.frame(canopy_layer = names(tukey_letters1$`canopy_layer`$Letters),
                        Letters = tukey_letters1$`canopy_layer`$Letters,
                        myc_type = rep("Total"))
tukey_df2 <- data.frame(var = names(tukey_letters2$`canopy_layer:myc_type`$Letters),
                        Letters = tukey_letters2$`canopy_layer:myc_type`$Letters)
tukey_df2$myc_type <- sub(".*:", "", tukey_df2$var)
tukey_df2$canopy_layer <- sub(":.*", "", tukey_df2$var)
tukey_df2 <- tukey_df2[,-1]
tukey_df <- rbind(tukey_df1, tukey_df2)

tropical_plot_data <- merge(tropical_plot_data, tukey_df, by = c("myc_type", "canopy_layer"), all = TRUE)
tropical_plot_data$canopy_layer <- as.numeric(tropical_plot_data$canopy_layer)

# Create plots
ggplot(tropical_plot_data, aes(x = canopy_layer, y = mean_simpson, color = myc_type)) +
  geom_point(size = 4) +
  geom_errorbar(aes(ymin = mean_simpson - se_simpson, ymax = mean_simpson + se_simpson), width = 0.2) +
  geom_line() +
  # geom_smooth(aes(group = myc_type), method = "lm", se = FALSE, linetype = "solid") +
  theme_minimal() +
  labs(
    title = "Simpson's diversity by Canopy Layer and Mycorrhizal Type (Tropical)",
    x = "Canopy Layer",
    y = "Mean Simpson's diversity (1-D)",
    color = "Mycorrhizal Type"
  ) +
  scale_color_manual(values = c("blue", "red", "black")) +
  geom_text(aes(label = Letters), vjust = 1.5, size = 7)

```



```
## Shannon diversity

# Load data
df <- read.csv("lambir_LDW.csv", stringsAsFactors = FALSE)

# Create a unique Site ID using site-specific columns
df$Site_ID <- paste(df$Forest, df$myc_type, df$plot_id, df$canopy_layer, sep = "-")

# Create a species-site abundance matrix
site_species_matrix <- df %>%
  group_by(Site_ID, Latin) %>%
  summarise(count = n(), .groups = "drop") %>%
  pivot_wider(names_from = Latin, values_from = count, values_fill = 0)

# Extract metadata and ensure species-site matrix is numeric
variable <- site_species_matrix[, 1, drop = FALSE] # Keep only Site_ID for merging later
site_species_matrix <- site_species_matrix[, -1] # Remove Site_ID column
site_species_matrix <- as.matrix(site_species_matrix) # Convert to matrix
mode(site_species_matrix) <- "numeric" # Ensure numeric format

rownames(site_species_matrix) <- unique(df$Site_ID) # Assign only unique Site_IDs

# Compute Shannon diversity
shannon <- diversity(site_species_matrix, index = "shannon")
```

```

# Store Shannon values in a dataframe
shan <- data.frame(Site_ID = rownames(site_species_matrix), Shannon = shannon)

variable <- df %>%
  distinct(Site_ID, .keep_all = TRUE) %>% # Ensure unique Site_IDs
  select(Site_ID, Forest, canopy_layer, myc_type) %>%
  left_join(shan, by = "Site_ID") %>%
  drop_na(Shannon) # Remove missing Shannon values

# Ensure structure is correct
str(variable)

## 'data.frame': 6773 obs. of 5 variables:
## $ Site_ID : chr "Tropical-EcM-0_0-1" "Tropical-AM-0_0-1" "Tropical-AM-0_0-2" "Tropical-EcM-0_0-1" ...
## $ Forest : chr "Tropical" "Tropical" "Tropical" "Tropical" ...
## $ canopy_layer: int 1 1 2 2 3 3 4 4 1 1 ...
## $ myc_type : chr "EcM" "AM" "AM" "EcM" ...
## $ Shannon : num 0.257 0.53 0.637 0.287 0 ...

## Function to summarize Shannon diversity
data_summary <- function(data) {
  data %>%
    group_by(canopy_layer, myc_type) %>%
    summarise(
      mean_Shannon = mean(Shannon, na.rm = TRUE),
      se_Shannon = sd(Shannon, na.rm = TRUE) / sqrt(n()),
      .groups = 'drop'
    )
}

## Function to add "Total" category
data_total <- function(data) {
  data %>%
    group_by(canopy_layer) %>%
    summarise(
      myc_type = "Total",
      mean_Shannon = mean(Shannon, na.rm = TRUE),
      se_Shannon = sd(Shannon, na.rm = TRUE) / sqrt(n()),
      .groups = 'drop'
    )
}

## Filter data for Temperate and Tropical forests
temperate_data <- variable %>% filter(Forest == "Temperate")
temperate_data$canopy_layer <- as.factor(temperate_data$canopy_layer)

# Run ANOVA - Total - Temperate
anova_result1 <- aov(Shannon ~ canopy_layer, data = temperate_data)
summary(anova_result1)

##           Df Sum Sq Mean Sq F value Pr(>F)
## canopy_layer    3    0.7  0.2192   0.584  0.626
## Residuals  1304  489.4  0.3753

```

```

tukey_result1 <- TukeyHSD(anova_result1)

# Run ANOVA - AM & EcM - Temperate
anova_result2 <- aov(Shannon ~ canopy_layer * myc_type, data = temperate_data)
summary(anova_result2)

##
##           Df Sum Sq Mean Sq F value Pr(>F)
## canopy_layer      3      0.7   0.2192   0.583  0.626
## myc_type          1      0.0   0.0129   0.034  0.853
## canopy_layer:myc_type  3      0.9   0.3025   0.805  0.491
## Residuals       1300    488.5   0.3758

tukey_result2 <- TukeyHSD(anova_result2)

## Process data for both forest types
temperate_plot_data <- bind_rows(data_summary(temperate_data), data_total(temperate_data))

# Create Tukey letters - Temperate
tukey_letters1 <- multcompLetters4(anova_result1, tukey_result1)
tukey_letters2 <- multcompLetters4(anova_result2, tukey_result2)
tukey_df1 <- data.frame(canopy_layer = names(tukey_letters1$`canopy_layer`$Letters),
                      Letters = tukey_letters1$`canopy_layer`$Letters,
                      myc_type = rep("Total"))
tukey_df2 <- data.frame(var = names(tukey_letters2$`canopy_layer:myc_type`$Letters),
                      Letters = tukey_letters2$`canopy_layer:myc_type`$Letters)
tukey_df2$myc_type <- sub(".*:", "", tukey_df2$var)
tukey_df2$canopy_layer <- sub(":.*", "", tukey_df2$var)
tukey_df2 <- tukey_df2[,-1]
tukey_df <- rbind(tukey_df1, tukey_df2)

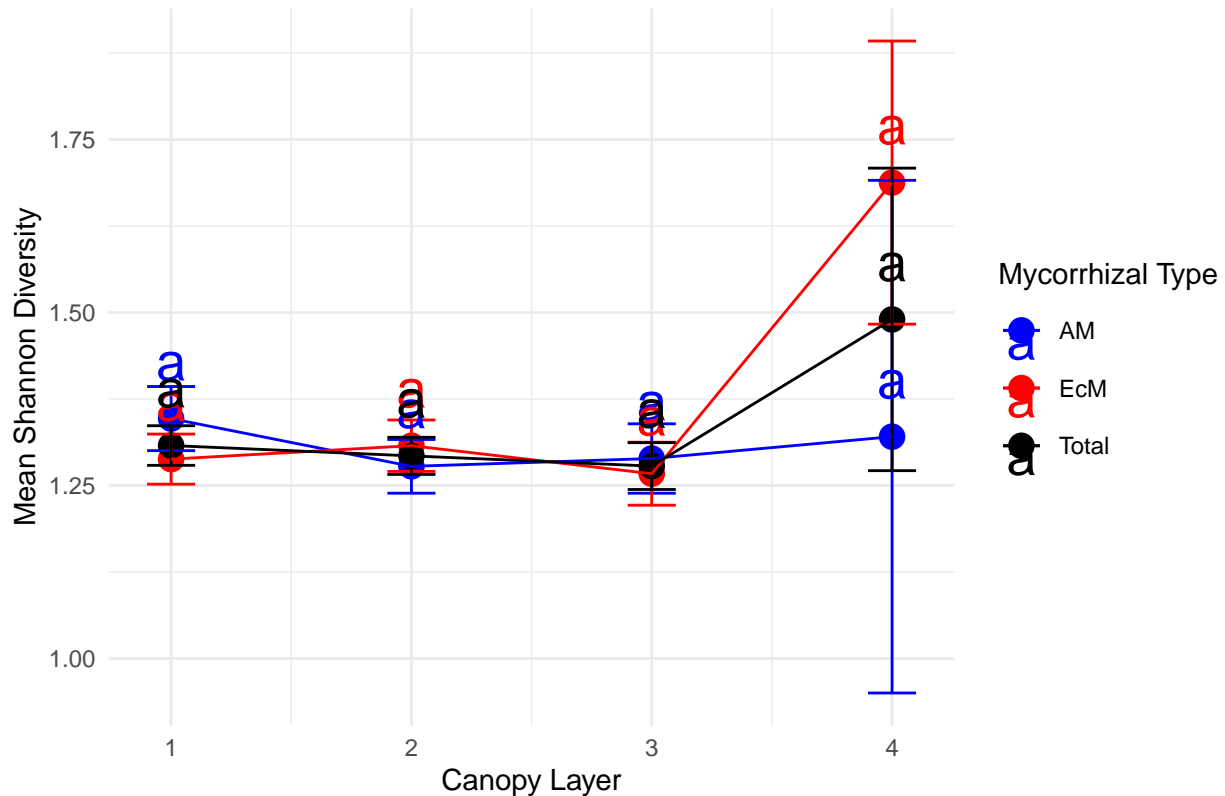
temperate_plot_data <- merge(temperate_plot_data, tukey_df, by = c("myc_type", "canopy_layer"), all = T)
temperate_plot_data$canopy_layer <- as.numeric(temperate_plot_data$canopy_layer)

## Create Shannon diversity plots

ggplot(temperate_plot_data, aes(x = canopy_layer, y = mean_Shannon, color = myc_type)) +
  geom_point(size = 4) +
  geom_errorbar(aes(ymin = mean_Shannon - se_Shannon, ymax = mean_Shannon + se_Shannon), width = 0.2) +
  geom_line() +
  theme_minimal() +
  labs(
    title = "Shannon Diversity by Canopy Layer and Mycorrhizal Type (Temperate)",
    x = "Canopy Layer",
    y = "Mean Shannon Diversity",
    color = "Mycorrhizal Type"
  ) +
  scale_color_manual(values = c("blue", "red", "black")) +
  geom_text(aes(label = Letters), vjust = -1, size = 7)

```

## Shannon Diversity by Canopy Layer and Mycorrhizal Type (Temperate)



```
# Repeat for tropical
tropical_data <- variable %>% filter(Forest == "Tropical")
tropical_plot_data <- bind_rows(data_summary(tropical_data), data_total(tropical_data))
tropical_data$canopy_layer <- as.factor(tropical_data$canopy_layer)
```

```
# Run ANOVA - Total - Tropical
anova_result1 <- aov(Shannon ~ canopy_layer, data = tropical_data)
summary(anova_result1)
```

```
##               Df Sum Sq Mean Sq F value Pr(>F)
## canopy_layer    6      36    6.05  2.654 0.0143 *
## Residuals   5458  12443    2.28
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
tukey_result1 <- TukeyHSD(anova_result1)
```

```
# Run ANOVA - AM & EcM - Tropical
anova_result2 <- aov(Shannon ~ canopy_layer * myc_type, data = tropical_data)
summary(anova_result2)
```

```
##               Df Sum Sq Mean Sq F value Pr(>F)
## canopy_layer    6      36    6.050  2.651 0.0144 *
## myc_type         1       0    0.160  0.070 0.7909
## canopy_layer:myc_type  6       3    0.492  0.216 0.9720
## Residuals       5451  12440    2.282
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```

tukey_result2 <- TukeyHSD(anova_result2)

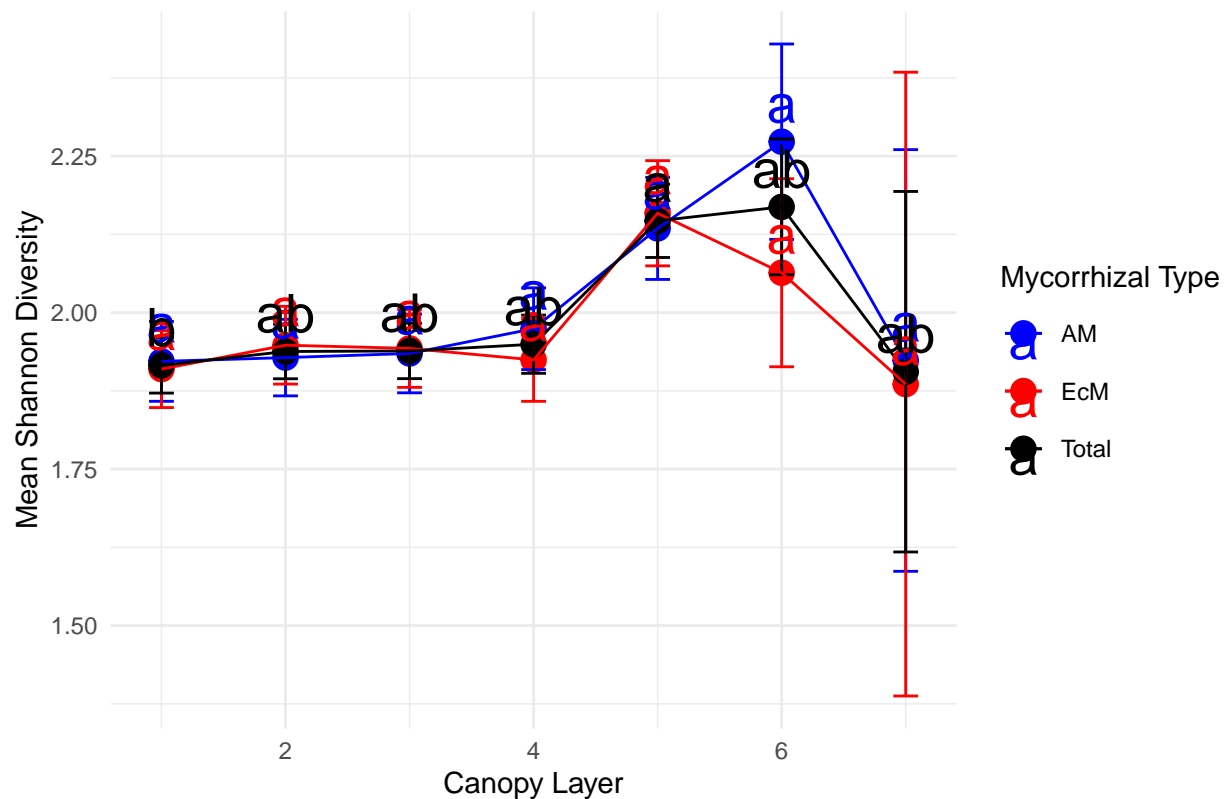
# Create Tukey letters - Tropical
tukey_letters1 <- multcompLetters4(anova_result1, tukey_result1)
tukey_letters2 <- multcompLetters4(anova_result2, tukey_result2)
tukey_df1 <- data.frame(canopy_layer = names(tukey_letters1$`canopy_layer`$Letters),
                        Letters = tukey_letters1$`canopy_layer`$Letters,
                        myc_type = rep("Total"))
tukey_df2 <- data.frame(var = names(tukey_letters2$`canopy_layer:myc_type`$Letters),
                        Letters = tukey_letters2$`canopy_layer:myc_type`$Letters)
tukey_df2$myc_type <- sub(".*:", "", tukey_df2$var)
tukey_df2$canopy_layer <- sub(":.*", "", tukey_df2$var)
tukey_df2 <- tukey_df2[,-1]
tukey_df <- rbind(tukey_df1, tukey_df2)

tropical_plot_data <- merge(tropical_plot_data, tukey_df, by = c("myc_type", "canopy_layer"), all = TRUE)
tropical_plot_data$canopy_layer <- as.numeric(tropical_plot_data$canopy_layer)

#Plot
ggplot(tropical_plot_data, aes(x = canopy_layer, y = mean_Shannon, color = myc_type)) +
  geom_point(size = 4) +
  geom_errorbar(aes(ymin = mean_Shannon - se_Shannon, ymax = mean_Shannon + se_Shannon), width = 0.2) +
  geom_line() +
  theme_minimal() +
  labs(
    title = "Shannon Diversity by Canopy Layer and Mycorrhizal Type (Tropical)",
    x = "Canopy Layer",
    y = "Mean Shannon Diversity",
    color = "Mycorrhizal Type"
  ) +
  scale_color_manual(values = c("blue", "red", "black")) +
  geom_text(aes(label = Letters), vjust = -0.5, size = 7)

```

## Shannon Diversity by Canopy Layer and Mycorrhizal Type (Tropical)



```
## Count occurrences of each species per site and get the top 2 frequent species
top_species_per_site <- df %>%
  group_by(Forest, myc_type, canopy_layer, Latin) %>%
  summarise(freq = n(), .groups = 'drop') %>%
  arrange(Forest, myc_type, canopy_layer, desc(freq)) %>%
  group_by(Forest, myc_type, canopy_layer) %>%
  slice_head(n = 2)

write.csv(top_species_per_site, "top2_species.csv")
```

### 3. Beta-Diversity - Hypothesis testing