Richness & Diversity

Jeff Oliver

September 20, 2018

Comparison of richness and diversity between collection methods

We are interested to compare Malaise trap and Pollard walk surveys to see if there is a difference in the observed species richness, R, and diversity. For the latter, we will use Shannon's H as a measure of diversity (Hill 1973).

Methods

Using the BioSCAN data, we first calculate R (richness) and H (diversity) for each site, including only those sites for which data from both survey types is available. For the latter, we use the `vegan` package in R (Oksanen et al. 2018). We then compare the two methods through paired t-tests, to determine if the difference in means between collection methods is significant. We use the `ggplot2` package (Wickham 2009) for visualization. Additional data-wrangling provided by `tidyr` package.

Starting by loading dependencies and reading in the data

```
library("tidyr")    # converting data to long and wide format
library("dplyr")    # data wrangling (group_by)
library("ggplot2")  # plotting
library("vegan")    # calculating Shannon's H
source(file = "bioscan-functions.R")
bioscan <- CompleteBioscan()

# Identify those columns with species data
species.cols <- c(5:33)
```

Richness

```
# Calculate richness for each row (total number of species with at least one
# individual observed)
bioscan$Richness <- apply(X = bioscan[, species.cols],
                          MARGIN = 1,
                          FUN = function(x) {
                            sum(x > 0)
                          })
```

Diversity

```
# Create data frame with only individual counts for each species.
bioscan$Diversity <- apply(X = bioscan[, species.cols],
                           MARGIN = 1,
                           FUN = function(x) {
                             vegan::diversity(x = x, index = "shannon")
                           })
```

t-test

Richness

```
# Create data frame for richness t-test. One column for Pollard, one for Malaise
richness.df <- bioscan[, c("Site.Number", "Collection.Method", "Richness")]
richness.df <- richness.df %>%
  spread(Collection.Method, Richness)

# Run the t-test
richness.t.test <- t.test(x = richness.df$Malaise,
                          y = richness.df$`Pollard Walk`,
                          paired = TRUE)

# Store the value of t
richness.t <- round(richness.t.test$statistic, 3)

# Rounding may turn p-values into zero, so account for that possibility here
richness.p <- "< 0.001"
if (richness.t.test$p.value > 0.001) {
  richness.p <- round(richness.t.test$p.value, 3)
}

# Store the estimated difference in means
richness.estimate <- round(richness.t.test$estimate, 3)

# Store the means
richness.means <- bioscan %>%
  group_by(Collection.Method) %>%
  summarise(means = round(mean(Richness), 3))
```

Diversity

```
# Create data frame for diversity t-test. One column for Pollard, one for Malaise
diversity.df <- bioscan[, c("Site.Number", "Collection.Method", "Diversity")]
diversity.df <- diversity.df %>%
  spread(Collection.Method, Diversity)

# Run the t-test
diversity.t.test <- t.test(x = diversity.df$Malaise,
                          y = diversity.df$`Pollard Walk`,
                          paired = TRUE)

# Store the value of t
diversity.t <- round(diversity.t.test$statistic, 3)

# Rounding may turn p-values into zero, so account for that possibility here
diversity.p <- "< 0.001"
if (diversity.t.test$p.value > 0.001) {
  diversity.p <- round(diversity.t.test$p.value, 3)
}

# Store the estimated difference in means
diversity.estimate <- round(diversity.t.test$estimate, 3)
```

```r
# Store the means
diversity.means <- bioscan %>%
  group_by(Collection.Method) %>%
  summarise(means = round(mean(Diversity), 3))
```

Results

Observed istribution of richness and diversity:

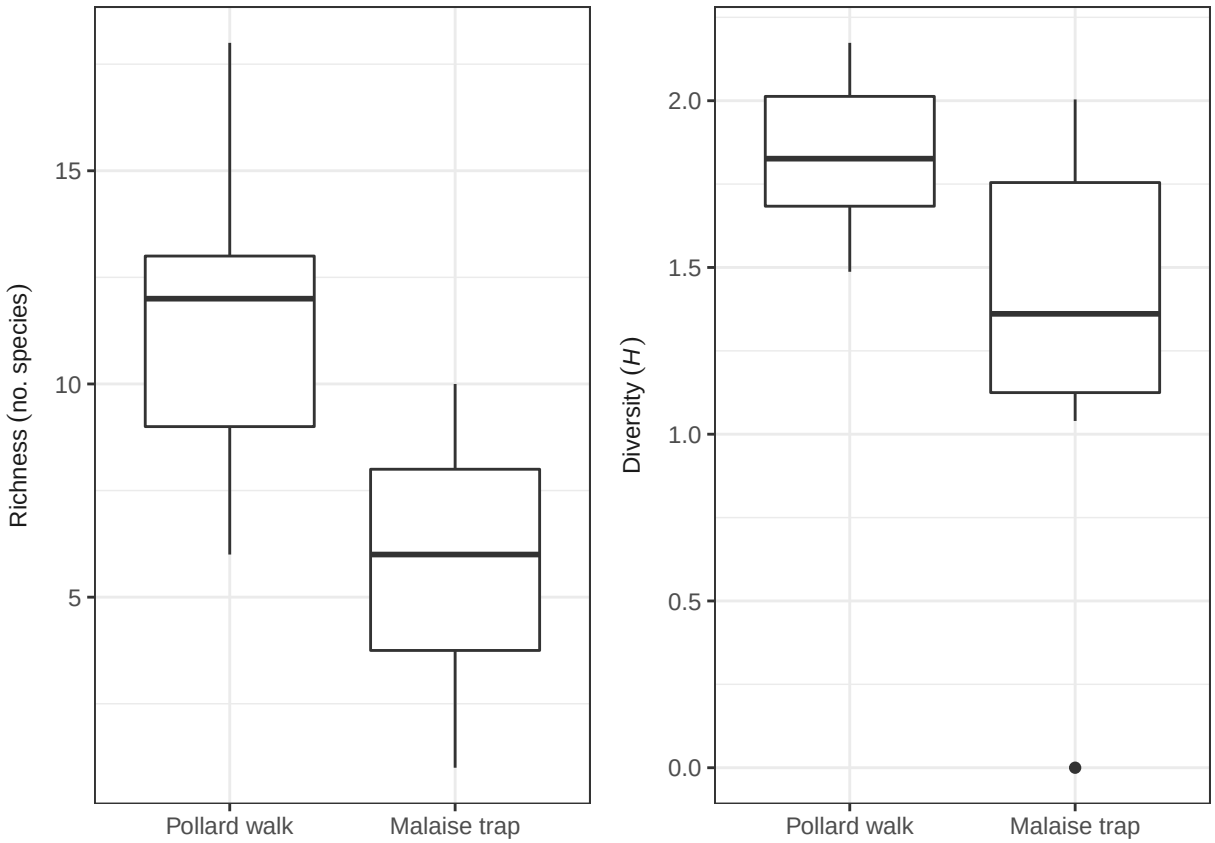
```r
# We need to create long-format data for use with ggplot
bioscan.long <- bioscan[, c("Collection.Method", "Richness", "Diversity")] %>%
  gather(key = "statistic",
         value = "value",
         -Collection.Method)

# For consistency, replace "Malaise" with "Malaise trap" and "Pollard Walk" with
# "Pollard walk"
bioscan.long$Collection.Method <- gsub(pattern = "Malaise",
                                       replacement = "Malaise trap",
                                       x = bioscan.long$Collection.Method)

bioscan.long$Collection.Method <- gsub(pattern = "Pollard Walk",
                                       replacement = "Pollard walk",
                                       x = bioscan.long$Collection.Method)


# Re-level the two factors so they appear in desired order
bioscan.long$Collection.Method <- factor(bioscan.long$Collection.Method,
                                         levels = c("Pollard walk", "Malaise trap"))
bioscan.long$statistic <- factor(bioscan.long$statistic,
                                 levels = c("Richness", "Diversity"))

# Boxplot, with a separate plot for each statistic
statistics.plot <- ggplot(data = bioscan.long,
                          mapping = aes(x = Collection.Method, y = value)) +
  geom_boxplot() +
  facet_wrap(~ statistic,
             scales = "free_y",
             strip.position = "left",
             labeller = as_labeller(c(Richness = "Richness~(no.~species)",
                                      Diversity = "Diversity~(italic(H))"),
                                    label_parsed)) +
  theme_bw() +
  ylab(NULL) +
  xlab(NULL) +
  theme(strip.background = element_blank(),
        strip.placement = "outside")
print(statistics.plot)
```




```
ggsave(filename = "output/richness-diversity-boxplot.png",
       plot = statistics.plot,
       width = 4.75,
       height = 3)
```

Means table:

Collection Method	R	H
Malaise trap	5.938	1.393
Pollard walk	11.312	1.846

t-table:

Statistic	t	p	δ
R	-4.657	< 0.001	-5.375
H	-3.307	0.005	-0.453