Ranas\_compare

S. Holden Jones

3/6/24

# COMPARE RICHNESS, ABUNDANCE, DIVERSITY, COMPOSITION ACROSS LANDUSE TYPES

This script will take processed data as input, run statistic tests and produce figures comparing the dependent variables (diversity, abundance, composition) of each landuse type broadly. Other scripts will explore the variation in these response variables w/in specific landuse types.

## ——————————————————————————–

load packages

load data ## weird that dfs are of different lengths, why are columns getting added randomly?. Also note that ranas is a tbl\_df and new is spec\_tbl\_df despite being created from ranas

define a consistent palet, labels, and theme to be used in multiple figures ## still would be nice to have palet for harvested comparison

# palet for comparisons of all 4 site types  
pal <- brewer.pal(4, "BrBG")  
  
# site labels for 4 site comparison, use in several plots  
site\_labels <- c("Shade \n (n = 12)", "Sun \n (n = 11)",   
 "Abandoned \n (n = 4)", "Forest \n (n = 5)")  
  
# site labels for harvest, non-harvest comparison  
labels\_harvest <- c("Not harvested \n (n = 9)", "Harvested \n (n = 23)")  
  
# set a theme for ggplots  
my\_theme <- function() {  
 theme\_minimal() +   
 theme(legend.position = "none",  
 plot.background = element\_rect("white"),  
 panel.background = element\_rect("white"),  
 panel.grid = element\_line("grey90"),  
 axis.line = element\_line("gray25"),  
 axis.text = element\_text(size = 12, color = "gray25"),  
 axis.title = element\_text(size = 14, color = "gray25"),  
 legend.text = element\_text(size = 12))  
}

## ——————————————————————————–

# RICHNESS ACROSS LANDUSE TYPES

richness by type - just number of species

## sig difference in species richness across landuse types

* although B, V are higher!

# group by site and type to retain type for ggplot  
richness\_by\_site <- new %>%  
 group\_by(Sitio, Tipo) %>%  
 summarize(species = list(sort(unique(`Final\_ID`))),  
 no\_species = n\_distinct(`Final\_ID`))

## `summarise()` has grouped output by 'Sitio'. You can override using the  
## `.groups` argument.

# Group by land\_use and calculate species richness  
richness\_summary <- richness\_by\_site %>%  
 group\_by(Tipo) %>%  
 summarize(mean\_species\_richness = mean(no\_species),  
 median\_species\_richness = median(no\_species),  
 min\_species\_richness = min(no\_species),  
 max\_species\_richness = max(no\_species))

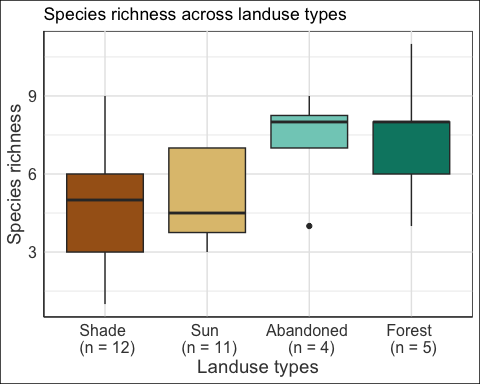
run ANOVA on species richness across the 4 landuse types - p-value: 0.0372

aov\_richness <- aov(no\_species ~ Tipo, data = richness\_by\_site)  
  
summary(aov\_richness)

## Df Sum Sq Mean Sq F value Pr(>F)   
## Tipo 3 42.83 14.278 3.205 0.0372 \*  
## Residuals 30 133.64 4.455   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

plot species richness across landuse types

# reorder factor levels  
richness\_by\_site$Tipo <- factor(richness\_by\_site$Tipo,  
 c("N", "C", "V", "B"))  
  
# plot species richness across landuse types using palet, my\_theme  
plot\_richness\_by\_site <- ggplot(richness\_by\_site,   
 aes(x = Tipo, y = no\_species, fill = Tipo)) +  
 geom\_boxplot() +  
 scale\_fill\_manual(values = pal) +  
 # order needs to match factor levels above  
 scale\_x\_discrete(labels = c(site\_labels)) +  
 my\_theme() +   
 labs(x = "Landuse types",  
 y = "Species richness",  
 title = "Species richness across landuse types")  
  
plot\_richness\_by\_site



|  |
| --- |
| # Harvested vs. non-harvested for richness |
| AOV sp. richness against harvested, non-harvested - highly sig! |
| ```r # add harvested column richness\_by\_site <- richness\_by\_site %>% mutate(harvested = case\_when( Tipo == “N” ~ “Yes”, Tipo == “C” ~ “Yes”, Tipo == “B” ~ “No”, Tipo == “V” ~ “No”) ) |
| # then run aov here aov\_richness\_harvest <- aov(no\_species ~ harvested, data = richness\_by\_site) |
| summary(aov\_richness\_harvest) ``` |
| ## Df Sum Sq Mean Sq F value Pr(>F) ## harvested 1 42.47 42.47 10.14 0.00322 \*\* ## Residuals 32 134.00 4.19 ## --- ## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1 |
| B+V have sig higher richness than N+C ## aesthetically would want to mix colors from previous plot |
| r # plot plot\_richness\_by\_harvest <- ggplot(richness\_by\_site, aes(x = harvested, y = no\_species, fill = harvested)) + scale\_x\_discrete(labels = c(labels\_harvest)) + geom\_boxplot() + my\_theme() + labs(x = "Active cacao plantation?", y = "Species richness") + ggtitle("Species richness in harvested and non-harvested areas") plot\_richness\_by\_harvest |
|  |

# ABUNDANCE ACROSS LANDUSE TYPES

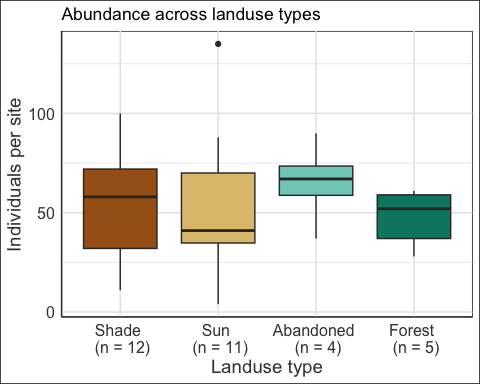
number of new individuals ## could eventually complicate this to look at pop size estimates?

anova for abundance - not sig

abundance\_by\_site <- new %>%  
 group\_by(Sitio, Tipo) %>%  
 count()  
  
abundance\_by\_site$Tipo <- factor(abundance\_by\_site$Tipo, c("N", "C", "V", "B"))  
  
aov\_abundance <- aov(n ~ Tipo, data = abundance\_by\_site)  
summary(aov\_abundance)

## Df Sum Sq Mean Sq F value Pr(>F)  
## Tipo 3 805 268.3 0.308 0.819  
## Residuals 30 26119 870.6

plot\_abundance\_by\_site <- ggplot(abundance\_by\_site, aes(x = Tipo, y = n, fill = Tipo)) +  
 geom\_boxplot() +  
 scale\_fill\_manual(values = pal) +  
 scale\_x\_discrete(labels = c(site\_labels)) +  
 my\_theme() +  
 labs(x = "Landuse type",  
 y = "Individuals per site",  
 title = "Abundance across landuse types")  
plot\_abundance\_by\_site



anova for abundance ## abundance differences are not statistically significant

aov\_abundance <- aov(n ~ Tipo, data = abundance\_by\_site)  
summary(aov\_abundance)

## Df Sum Sq Mean Sq F value Pr(>F)  
## Tipo 3 805 268.3 0.308 0.819  
## Residuals 30 26119 870.6

## ——————————————————————————–

# DIVERSITY ACROSS LANDUSE TYPES

Look at how different diversity metrics compare across landuse types

# First need to convert data into 3 column format

habitat, species, abundance for labdsv - use matrify!

new\_mat - three column format to use in labdsv package

three\_column <- new %>%  
 group\_by(Sitio) %>%  
 count(`Final\_ID`)  
  
# needs to be converted to data.frame!  
three\_column <- data.frame(three\_column)  
  
new\_mat <- matrify(three\_column)

| # Diversity Indices |
| --- |
| # dataframe creation here to run statistical tests |
| create site\_type df and join to enframed shannon\_df |
| ```r # create site\_type df, just name of site and the type site\_type <- site\_data %>% filter(Transecto == 1) %>% # so just one row per site select(Sitio, Tipo) |
| # reorder factor levels now for downstream plotting site\_typeTipo, c(“N”, “C”, “V”, “B”)) |
| # here add harvested column for later analyses site\_type <- site\_type %>% mutate(harvested = case\_when( Tipo == “N” ~ “Yes”, Tipo == “C” ~ “Yes”, Tipo == “B” ~ “No”, Tipo == “V” ~ “No”) ) |
| # enframe shannon which is currently a vector shannon\_df <- enframe(shannon) |
| # need to change name to Sitio in shannon\_df to match for left\_join # also change value to Shannon\_Index colnames(shannon\_df) <- c(“Sitio”,“Shannon\_Index”) |
| # join by “Sitio” shannon\_df <- left\_join(site\_type, shannon\_df, by = “Sitio”) ``` |
| do the same thing for simpson\_df |
| ```r simpson\_df <- enframe(simpson) |
| colnames(simpson\_df) <- c(“Sitio”,“Simpson\_Index”) |
| simpson\_df <- left\_join(site\_type, simpson\_df, by = “Sitio”) ``` |
| and inverse simpson\_df |
| ```r inv\_simpson\_df <- enframe(inv\_simpson) |
| colnames(inv\_simpson\_df) <- c(“Sitio”,“Inv\_Simpson\_Index”) |
| inv\_simpson\_df <- left\_join(site\_type, inv\_simpson\_df, by = “Sitio”) ``` |

# run aovs across landuse types for each diversity index

none of these are significant currently, but shannon and simpson are close - as more sequencing results come back it’s possible this will become significant!

shannon\_aov <- aov(Shannon\_Index ~ Tipo, data = shannon\_df)  
summary(shannon\_aov)

## Df Sum Sq Mean Sq F value Pr(>F)  
## Tipo 3 0.913 0.3043 1.976 0.14  
## Residuals 28 4.312 0.1540

# not sig - p = 0.14!  
  
simpson\_aov <- aov(Simpson\_Index ~ Tipo, data = simpson\_df)  
summary(simpson\_aov)

## Df Sum Sq Mean Sq F value Pr(>F)  
## Tipo 3 0.1794 0.05979 1.638 0.203  
## Residuals 28 1.0218 0.03649

# not sig - p = 0.2!  
  
inv\_simpson\_aov <- aov(Inv\_Simpson\_Index ~ Tipo, data = inv\_simpson\_df)  
summary(inv\_simpson\_aov)

## Df Sum Sq Mean Sq F value Pr(>F)  
## Tipo 3 1.239 0.4129 0.837 0.485  
## Residuals 28 13.806 0.4931

# not sig - p = 0.49!

| # plot how diversity indices vary across land types - these are pretty much identical regardless of index used |
| --- |
| # Compare diversity indices for harvested vs. unharvested sites |
| do the same thing as above, using harvested, unharvested column - aovs across harvested, unharvested |
| r # shannon - sig shannon\_harvest\_aov <- aov(Shannon\_Index ~ harvested, data = shannon\_df) summary(shannon\_harvest\_aov) |
| ## Df Sum Sq Mean Sq F value Pr(>F) ## harvested 1 0.810 0.8096 5.501 0.0258 \* ## Residuals 30 4.415 0.1472 ## --- ## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1 |
| r # simpson - sig simpson\_harvest\_aov <- aov(Simpson\_Index ~ harvested, data = simpson\_df) summary(simpson\_harvest\_aov) |
| ## Df Sum Sq Mean Sq F value Pr(>F) ## harvested 1 0.1513 0.1513 4.322 0.0463 \* ## Residuals 30 1.0499 0.0350 ## --- ## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1 |
| r # inverse simpson - not sig inv\_simpson\_harvest\_aov <- aov(Inv\_Simpson\_Index ~ harvested, data = inv\_simpson\_df) summary(inv\_simpson\_harvest\_aov) |
| ## Df Sum Sq Mean Sq F value Pr(>F) ## harvested 1 1.027 1.0266 2.197 0.149 ## Residuals 30 14.018 0.4673 |
| now make figures for each ## aesthetically think about fill colors |
| shannon\_harvest |
| r plot\_shannon\_harvest <- ggplot(shannon\_df, aes(x = harvested, y = Shannon\_Index, fill = harvested)) + geom\_boxplot() + my\_theme() + labs(x = "Harvested", y = "Shannon diversity", title = "Shannon diversity across harvested and unharvested sites") plot\_shannon\_harvest |
|  |
| simpson\_harvest |
| r plot\_simpson\_harvest <- ggplot(simpson\_df, aes(x = harvested, y = Simpson\_Index, fill = harvested)) + geom\_boxplot() + my\_theme() + labs(x = "Harvested", y = "Simpson diversity", title = "Simpson diversity across harvested and unharvested sites") plot\_simpson\_harvest |
|  |
| inv\_simpson\_harvest |
| r plot\_inv\_simpson\_harvest <- ggplot(inv\_simpson\_df, aes(x = harvested, y = Inv\_Simpson\_Index, fill = harvested)) + geom\_boxplot() + my\_theme() + labs(x = "Harvested", y = "Inverse Simpson diversity", title = "Inverse Simpson diversity across harvested and unharvested sites") plot\_inv\_simpson\_harvest |
|  |
| ## AND RUN FALSE DISCOVERY RATE? A POST-HOC TEST NEEDED FOR DIVERSITY INDICES |

# COMPOSITION ACROSS LANDUSE TYPES….

| # Pair-wise dissimilarity |
| --- |
| # Non-metric Multidimensional Scaling |
| ## need to reorder new\_mat rows to allow nmds to draw borders of communities - goal is to represent info from multiple dimensions into a few, so can visualize and interpret. - so here, present the similarity of communities in 2-dimensional space |
| this is going to be ugly, but have triple checked against original new\_mat |
| ```r # Define the desired order of rows: Shade, Sun, Abandoned, Forest new\_order <- c(1,3,5,8,19,20,22,25,27,29,30,32, # shade (n=12) 2,4,7,16,18,21,23,24,26,28,31, # sun (n=11) 9,13,14,17, # abandoned (n=4) 6,10,11,12,15 # forest (n=5) ) |
| # Reorder rows new\_mat <- new\_mat[new\_order, ] ``` |

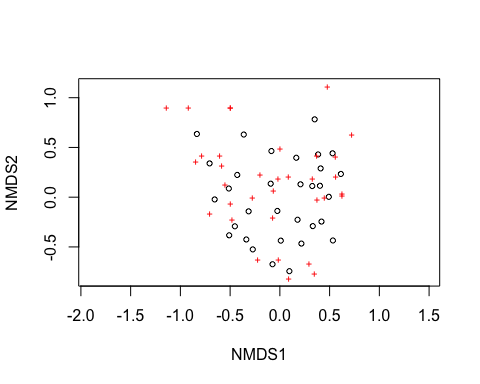
## create NMDS plot and draw boundaries of landuse types

code from Peat Clark’s vegan tutorial

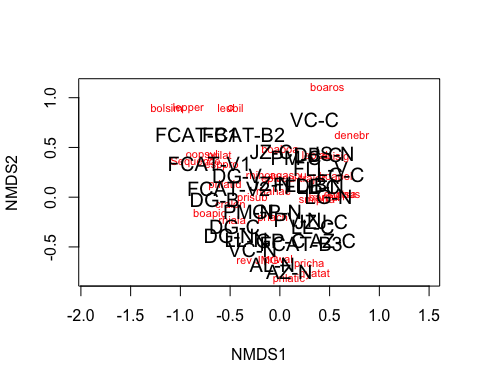
# use community by species matrix from above  
new\_mat\_NMDS=metaMDS(new\_mat, k=2)

## Square root transformation  
## Wisconsin double standardization  
## Run 0 stress 0.2494371   
## Run 1 stress 0.2494872   
## ... Procrustes: rmse 0.01160053 max resid 0.04667101   
## Run 2 stress 0.3005787   
## Run 3 stress 0.2550878   
## Run 4 stress 0.2600845   
## Run 5 stress 0.2494871   
## ... Procrustes: rmse 0.01165815 max resid 0.04696473   
## Run 6 stress 0.2627076   
## Run 7 stress 0.258064   
## Run 8 stress 0.2623353   
## Run 9 stress 0.2910167   
## Run 10 stress 0.2810836   
## Run 11 stress 0.2494872   
## ... Procrustes: rmse 0.0116367 max resid 0.04684101   
## Run 12 stress 0.2576727   
## Run 13 stress 0.2580647   
## Run 14 stress 0.249437   
## ... New best solution  
## ... Procrustes: rmse 3.930045e-05 max resid 0.0001178085   
## ... Similar to previous best  
## Run 15 stress 0.2630216   
## Run 16 stress 0.2576727   
## Run 17 stress 0.2542954   
## Run 18 stress 0.2494874   
## ... Procrustes: rmse 0.01175254 max resid 0.04737963   
## Run 19 stress 0.2516927   
## Run 20 stress 0.2577904   
## \*\*\* Best solution repeated 1 times

plot(new\_mat\_NMDS)



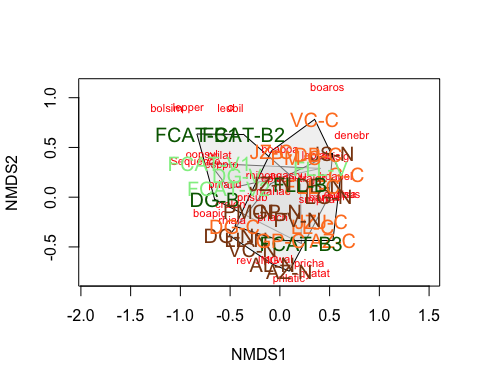
# here sites are open circles, species are red crosses  
  
# here help visualize by labeling specific sites and species  
  
# ordination plot function especially for congested plots  
ordiplot(new\_mat\_NMDS,type="n")  
# this function adds text or points to ordination plots  
orditorp(new\_mat\_NMDS,display="species",col="red",air=0.01)  
  
orditorp(new\_mat\_NMDS,display="sites",cex=1.25,air=0.01)



## use ordihull to help visualize different treatments unifying sites

assign treatment levels per landuse type on reordered new\_mat\_reordered

# ensure that labels are in correct order and have correct length  
landuse\_treatments=c(rep("Shade",12),  
 rep("Sun",11),   
 rep("Abandoned",4),   
 rep("Forest",5))  
  
ordiplot(new\_mat\_NMDS,type="n")  
  
ordihull(new\_mat\_NMDS,groups=landuse\_treatments,draw="polygon",  
 col="grey90",label=F)  
  
orditorp(new\_mat\_NMDS,display="species",col="red",air=0.01)  
  
# here colors need to align with landuse\_treatments from above!  
# of course need to use chocolate!  
orditorp(new\_mat\_NMDS,display="sites",col=c(rep("chocolate4",12),  
 rep("chocolate1",11),  
 rep("lightgreen",4),  
 rep("darkgreen",5)),   
 air=0.01,cex=1.25)



## very cool! see a lot of overlap in species simmilarity across landuse types

* but, forest and abandoned sites are similar, minus 3 outliers (FLL-B, FLL-C, FCAT-B3)
* these are also the lower elevation sites of the 9!
* think that elevation will be an important thing to include when comparing communities! ## CONTINUE HERE looking at elevation and group identity for NMDS

## ——————————————————————————–

# RAREFACTION….This is more of a richness approach?

## COPIED FROM RANAS\_EXPLORE \_ REVIEW

technique to assess expected species richness - allows calculation of species richness for given number of samples based on construction of rarefaction curves

issue is that the larger number of ind sampled, more species will be found - rarefaction curves created by randomly re-sampling pool of N samples multiple times and then plotting average number of species found in each sample - rarefaction generates expected number of species in small collection of n ind drawn from large pool of N samples - typically grow rapidly at first, then slowly as only rarest species remain to be sampled

use rarefy and rarecurve functions

spAbund <- rowSums(new\_mat)  
spAbund

## AL-N AZ-N DB-N DG-N GP-N JS-N JZ-N LG-N LL-N PM-N   
## 32 27 93 100 39 58 30 71 59 53   
## PV-N VC-N AZ-C DB-C DG-C FLL-C GP-C JZ-C JZJ-C LG-C   
## 77 70 22 60 135 88 40 42 79 50   
## LL-C PM-C VC-C DG-V FCAT-V1 FCAT-V2 FLL-V DG-B FCAT-B1 FCAT-B2   
## 40 67 6 66 68 37 90 52 37 59   
## FCAT-B3 FLL-B   
## 58 28

test <- colSums(new\_mat)  
test

## agaspu barpul boaboa boapel boapic boaros bolsim cralon   
## 5 70 3 67 3 1 1 7   
## denebr engpus epibou epiesp esppro hyatat leplab lepper   
## 1 4 85 1 1 3 49 3   
## leubil oopsyl priach pricha prilat prilatic prilatid prisub   
## 1 20 1389 1 8 2 8 17   
## priwal rev JMG rhahae rhiala rhihor scisig scitsa Sequence   
## 10 3 3 19 3 1 19 4   
## smipha trajor   
## 5 16

rarefaction uses smallest number of obs per sample to extrapolate expected number if all other samples only had that number of obs - VC-C only had 3!!! may want to proceed without this in the future - GP-N also only 9, without these two the next lowest is 22

# including VC-C raremin is 3  
raremin <- min(rowSums(new\_mat))  
raremin

## [1] 6

sRare <- rarefy(new\_mat, raremin)  
sRare

## AL-N AZ-N DB-N DG-N GP-N JS-N JZ-N LG-N   
## 1.750000 1.623932 3.018008 1.176970 1.748988 3.367215 2.296540 2.339415   
## LL-N PM-N PV-N VC-N AZ-C DB-C DG-C FLL-C   
## 1.279400 1.861018 1.666496 1.250932 2.025974 3.267734 1.256915 2.392368   
## GP-C JZ-C JZJ-C LG-C LL-C PM-C VC-C DG-V   
## 1.942547 2.092329 1.365394 2.430235 1.430769 2.232814 3.000000 2.210769   
## FCAT-V1 FCAT-V2 FLL-V DG-B FCAT-B1 FCAT-B2 FCAT-B3 FLL-B   
## 2.280941 1.745817 3.170037 2.981942 3.172652 2.689477 1.766622 2.745835   
## attr(,"Subsample")  
## [1] 6

# it won't allow us to override this to 22 (makes sense)  
# so will need to remove VC-C, GP-N from sites, via row position  
new\_mat\_abundant <- new\_mat %>%  
 filter(!row\_number() %in% c(19, 31))  
  
# let's see what happens with this as our raremin, and only using 30/32 sites  
raremin\_abundant <- min(rowSums(new\_mat\_abundant))  
raremin\_abundant

## [1] 6

# sRare gives expected 'rarefied' number of species (not obs) if only 22 collected  
sRare\_abundant <- rarefy(new\_mat\_abundant, raremin\_abundant)  
sRare\_abundant

## AL-N AZ-N DB-N DG-N GP-N JS-N JZ-N LG-N   
## 1.750000 1.623932 3.018008 1.176970 1.748988 3.367215 2.296540 2.339415   
## LL-N PM-N PV-N VC-N AZ-C DB-C DG-C FLL-C   
## 1.279400 1.861018 1.666496 1.250932 2.025974 3.267734 1.256915 2.392368   
## GP-C JZ-C LG-C LL-C PM-C VC-C DG-V FCAT-V1   
## 1.942547 2.092329 2.430235 1.430769 2.232814 3.000000 2.210769 2.280941   
## FCAT-V2 FLL-V DG-B FCAT-B1 FCAT-B2 FLL-B   
## 1.745817 3.170037 2.981942 3.172652 2.689477 2.745835   
## attr(,"Subsample")  
## [1] 6

shannon\_aov <- aov(shannon ~ Tipo, data = site\_type)  
summary(shannon\_aov)

## Df Sum Sq Mean Sq F value Pr(>F)  
## Tipo 3 0.221 0.0738 0.413 0.745  
## Residuals 28 5.003 0.1787

## no statistical difference in rarefied species richness across types

*note that using rarefaction to remove effects of diff sample sizes is bad!*

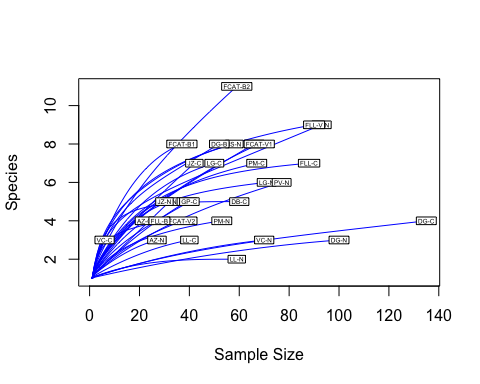
# join tipo data with sRare\_abundant in a df  
sRare\_abundant\_df <- sRare\_abundant %>%   
 enframe() %>%   
 full\_join(site\_type, by = c("name" = "Sitio"))  
  
sRare\_abundant\_aov <- aov(value ~ Tipo, data = sRare\_abundant\_df)  
summary(sRare\_abundant\_aov)

## Df Sum Sq Mean Sq F value Pr(>F)   
## Tipo 3 2.793 0.9309 2.444 0.0866 .  
## Residuals 26 9.903 0.3809   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
## 2 observations deleted due to missingness

# p-value is 0.183 (04/9/23) - check back in when sequencing is done!

visualization using rarecurve, for ggplot? ## make rarecurve for each type of site ## stuck!

# rarecurve for each site, would be nice to have for each type!  
rarecurve(new\_mat\_abundant, col = "blue", cex = 0.4)



# TEST - try to make separate curve for each type  
# make dfs of equal length  
#site\_type\_abundant <- site\_type %>%  
# filter(!Sitio == 'VC-C',  
# !Sitio == 'GP-N')  
  
  
# may need these as vector eventually?  
#test <- cbind(new\_mat\_abundant, site\_type\_abundant)  
  
#rarecurve\_test <- rarecurve(test, step=1, label=TRUE, col = test$Tipo,   
# xlab = "Number of individuals sampled", ylab = "Species richness")