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))

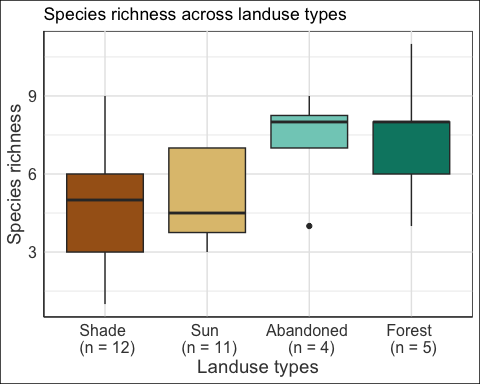
run ANOVA on species richness across the 4 landuse types - p-value: 0.0372

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

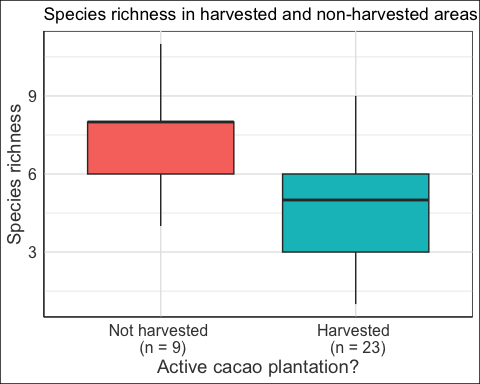
AOV sp. richness against harvested, non-harvested - highly sig!

richness\_by\_site <- richness\_by\_site %>%  
 mutate(harvested = case\_when( # add harvested column  
 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

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



## rarified species richness

# NEED TO GO THROUGH rarefy\_scrap to piece this piece together

## ——————————————————————————–

# ABUNDANCE ACROSS LANDUSE TYPES

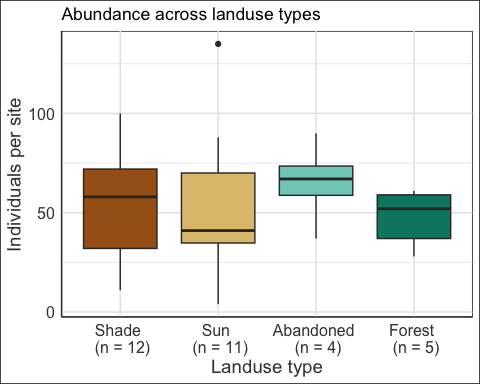
number of new individuals ## could eventually complicate this to look at pop size estimates?

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

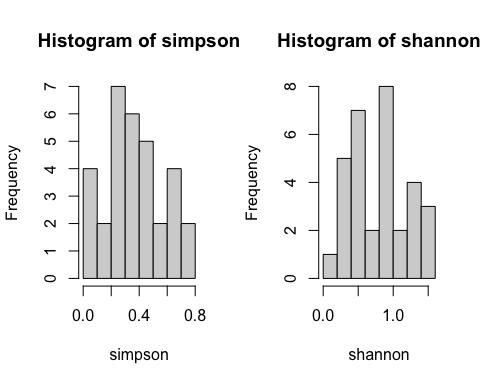
<https://peat-clark.github.io/BIO381/veganTutorial.html> <https://www.youtube.com/watch?v=wq1SXGQYgCs>

use 3 column format to get diversity indices for each site!

shannon <- diversity(new\_mat, index = "shannon")  
simpson <- diversity(new\_mat, index = "simpson")  
inv\_simpson <- diversity(new\_mat, index = "invsimpson")

compare these indices with histograms

par(mfrow = c(1, 2))   
# use par to generate panels with 1 row of 2 graphs  
hist(simpson)  
hist(shannon)



## dataframe creation here to run statistical tests

create site\_type df and join to enframed shannon\_df

site\_type <- site\_data %>% # create site\_type df, just name of site and the type  
 filter(Transecto == 1) %>% # so just one row per site  
 select(Sitio, Tipo)  
  
# reorder factor levels now for downstream plotting  
site\_type$Tipo <- factor(site\_type$Tipo, 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

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

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

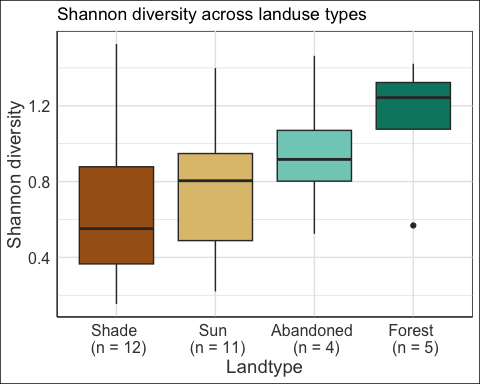
# not sig - p = 0.49!

## plot how diversity indices vary across land types

* these are pretty much identical regardless of index used

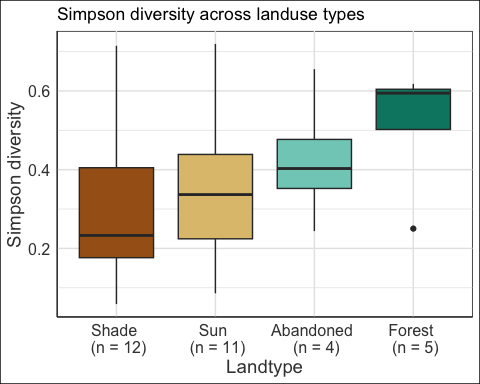
shannon

plot\_shannon <- ggplot(shannon\_df,   
 aes(x = Tipo, y = Shannon\_Index, fill = Tipo)) +  
 geom\_boxplot() +  
 scale\_fill\_manual(values = pal) +  
 scale\_x\_discrete(labels = c(site\_labels)) +  
 my\_theme() +  
 labs(x = "Landtype",  
 y = "Shannon diversity",  
 title = "Shannon diversity across landuse types")  
plot\_shannon



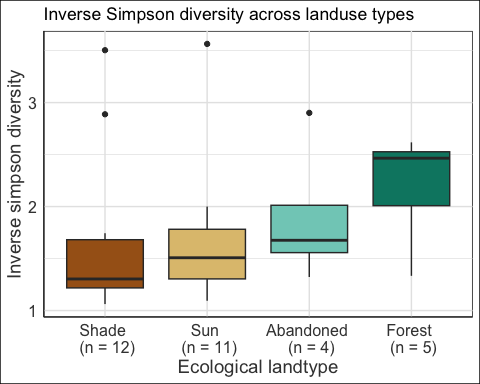
simpson

plot\_simpson <- ggplot(simpson\_df,   
 aes(x = Tipo, y = Simpson\_Index, fill = Tipo)) +  
 geom\_boxplot() +  
 scale\_fill\_manual(values = pal) +  
 scale\_x\_discrete(labels = c(site\_labels)) +  
 my\_theme() +   
 labs(x = "Landtype",  
 y = "Simpson diversity",  
 title = "Simpson diversity across landuse types")  
plot\_simpson



inv simpson

plot\_inv\_simpson <- ggplot(inv\_simpson\_df,   
 aes(x = Tipo, y = Inv\_Simpson\_Index, fill = Tipo)) +  
 geom\_boxplot() +  
 scale\_fill\_manual(values = pal) +  
 scale\_x\_discrete(labels = c(site\_labels)) +  
 my\_theme() +  
 labs(x = "Ecological landtype",  
 y = "Inverse simpson diversity",  
 title = "Inverse Simpson diversity across landuse types")  
plot\_inv\_simpson



## Compare diversity indices for harvested vs. unharvested sites

do the same thing as above, using harvested, unharvested column - aovs across harvested, unharvested

shannon\_harvest\_aov <- aov(Shannon\_Index ~ harvested, data = shannon\_df)  
summary(shannon\_harvest\_aov) # shannon - sig

## 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

simpson\_harvest\_aov <- aov(Simpson\_Index ~ harvested, data = simpson\_df)  
summary(simpson\_harvest\_aov) # simpson - sig

## 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

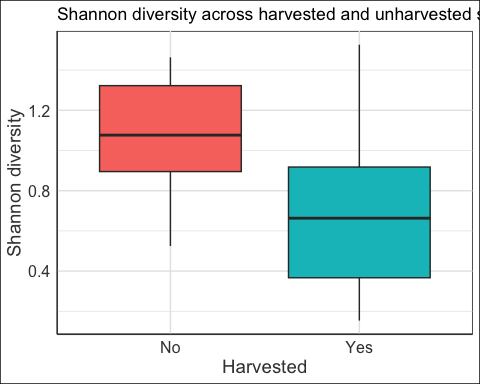
inv\_simpson\_harvest\_aov <- aov(Inv\_Simpson\_Index ~ harvested, data = inv\_simpson\_df)  
summary(inv\_simpson\_harvest\_aov) # inverse simpson - not sig

## 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

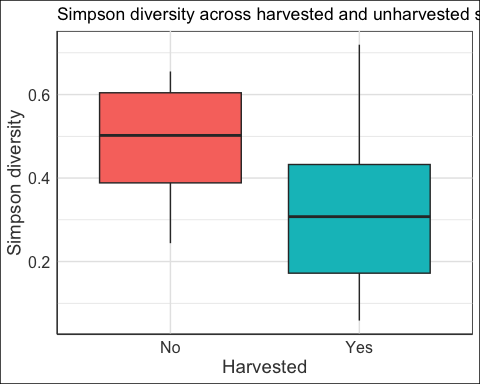
shannon\_harvest

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



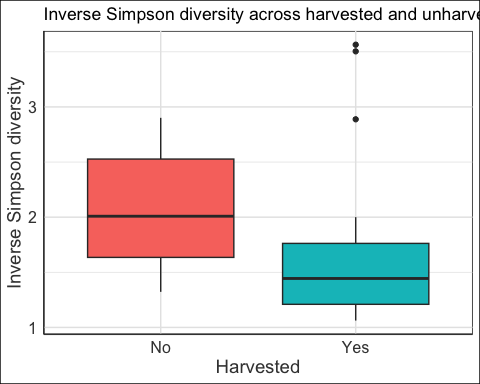
simpson\_harvest

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

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



## NEXT STEP: RUN A POST-HOC TEST NEEDED FOR DIVERSITY INDICES

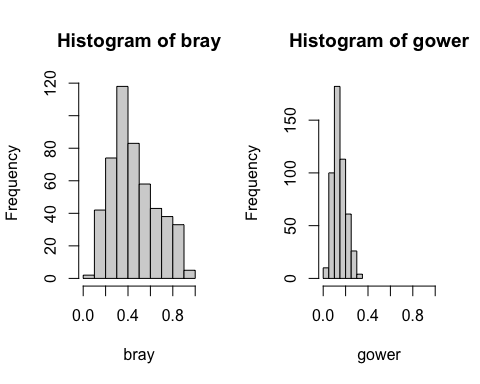
## ——————————————————————————–

# COMPOSITION ACROSS LANDUSE TYPES….

## Pair-wise dissimilarity

Calculate pair-wise dissimilarity (distance) using vegdist - gower and bray-curtis are good in detecting underlying ecological gradients

par(mfrow = c(1, 2))  
bray <- vegdist(new\_mat, "bray")   
gower <- vegdist(new\_mat, "gower")  
  
hist(bray, xlim = range(0.0,1.0))  
hist(gower, xlim = range(0.0,1.0))



## CONTINUE HERE

dissimilarity analysis is good way to explore variability in community comp - next steps would be to do some sort of cluster analysis - see where community associations exist - but switching gears to look at rarefaction

## 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

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)  
 )  
# Defined the desired order of rows: Shade, Sun, Abandoned, Forest  
  
# Reorder rows  
new\_mat <- new\_mat[new\_order, ]

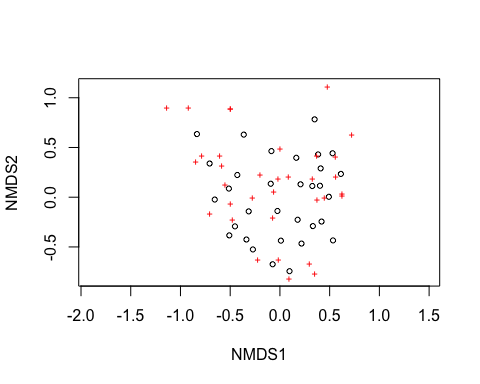
## create NMDS plot and draw boundaries of landuse types

code from Peat Clark’s vegan tutorial

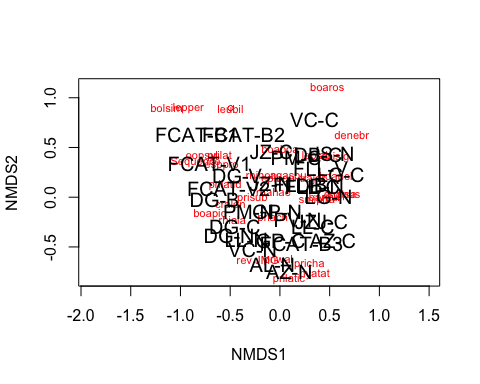
new\_mat\_NMDS=metaMDS(new\_mat, k=2)

## Square root transformation  
## Wisconsin double standardization  
## Run 0 stress 0.2494371   
## Run 1 stress 0.2577921   
## Run 2 stress 0.2552666   
## Run 3 stress 0.2494874   
## ... Procrustes: rmse 0.01160559 max resid 0.04669039   
## Run 4 stress 0.2494874   
## ... Procrustes: rmse 0.0115929 max resid 0.04663612   
## Run 5 stress 0.2546979   
## Run 6 stress 0.2494874   
## ... Procrustes: rmse 0.01158939 max resid 0.04661991   
## Run 7 stress 0.2550878   
## Run 8 stress 0.2545889   
## Run 9 stress 0.2494871   
## ... Procrustes: rmse 0.01165082 max resid 0.04690347   
## Run 10 stress 0.2610522   
## Run 11 stress 0.2494873   
## ... Procrustes: rmse 0.01161124 max resid 0.04672541   
## Run 12 stress 0.2494871   
## ... Procrustes: rmse 0.01164145 max resid 0.04686519   
## Run 13 stress 0.2494871   
## ... Procrustes: rmse 0.0116641 max resid 0.04695892   
## Run 14 stress 0.2494371   
## ... Procrustes: rmse 0.0002306455 max resid 0.0009847854   
## ... Similar to previous best  
## Run 15 stress 0.2841221   
## Run 16 stress 0.2554136   
## Run 17 stress 0.2494371   
## ... Procrustes: rmse 0.0003664102 max resid 0.001552107   
## ... Similar to previous best  
## Run 18 stress 0.2494872   
## ... Procrustes: rmse 0.01163332 max resid 0.04683154   
## Run 19 stress 0.2516926   
## Run 20 stress 0.2797229   
## \*\*\* Best solution repeated 2 times

# use community by species matrix from above  
  
plot(new\_mat\_NMDS)



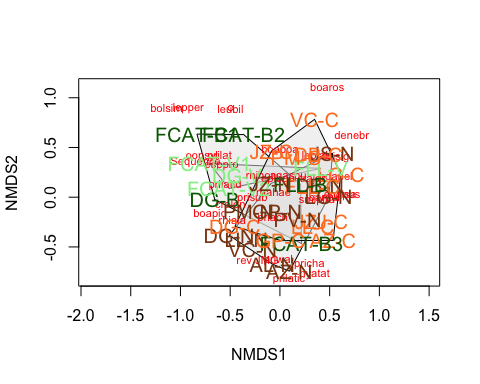
# here sites are open circles, species are red crosses  
  
# here help visualize by labeling specific sites and species  
  
# ordination plot function especially for congested plots  
ordiplot(new\_mat\_NMDS,type="n")  
# this function adds text or points to ordination plots  
orditorp(new\_mat\_NMDS,display="species",col="red",air=0.01)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

landuse\_treatments=c(rep("Shade",12),  
 rep("Sun",11),   
 rep("Abandoned",4),   
 rep("Forest",5))  
# ensure that labels are in correct order and have correct length  
  
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

## NEXT STEPS: MAKE ORDINATION OFF GROUP, COMMUNITY DISSIMILARITY

## ——————————————————————————–