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

## LOOK HERE TO REVIEW IMPORT OF NEW\_MAT - MAYBE DO MATRIX INSTEAD OF TIBBLE?

load data

define a consistent palet, labels, and theme to be used in 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))  
}

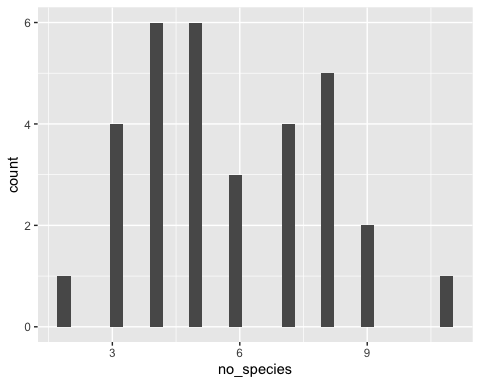
# RICHNESS ACROSS LANDUSE TYPES

## before selecting statistical test, need to look at dist. of richness by site

* use Poisson dist. b/c using count data with no upper limit!

(richness\_by\_site\_hist <- ggplot(richness\_by\_site, aes(no\_species)) +  
 geom\_histogram())

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



## run glm for species richness across landuse types, using Poisson dist.

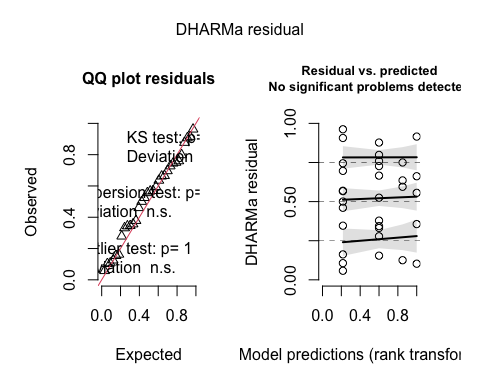
## NO SIG DIFF IN RICHNESS ACROSS LANDUSE TYPES

* bosque is sig diff from nacional w/out elevation as random fact - viejo almost
* bosque and viejo very close to being significant - need final IDs back!
* using poisson dist in glm - DHARMa simulation shows that it works well

richness\_by\_type\_glm <- glm(no\_species ~ Tipo + (1|Elevation),   
 data = richness\_by\_site,   
 family = poisson(link = "log"))  
summary(richness\_by\_type\_glm)

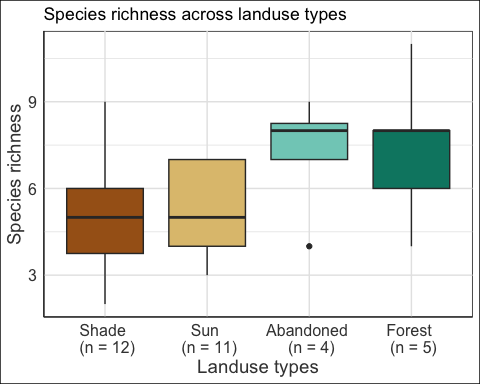
##   
## Call:  
## glm(formula = no\_species ~ Tipo + (1 | Elevation), family = poisson(link = "log"),   
## data = richness\_by\_site)  
##   
## Coefficients: (1 not defined because of singularities)  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 2.00148 0.16440 12.175 <2e-16 \*\*\*  
## TipoC -0.37402 0.21186 -1.765 0.0775 .   
## TipoN -0.37551 0.20838 -1.802 0.0715 .   
## TipoV -0.02048 0.24801 -0.083 0.9342   
## 1 | ElevationTRUE NA NA NA NA   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for poisson family taken to be 1)  
##   
## Null deviance: 25.719 on 31 degrees of freedom  
## Residual deviance: 20.280 on 28 degrees of freedom  
## AIC: 141.59  
##   
## Number of Fisher Scoring iterations: 4

# here use DHARMa package to see if we're using the correct model  
# no significant problems detected!  
simulationOutput\_richness\_by\_type <- simulateResiduals(  
 fittedModel = richness\_by\_type\_glm)  
plot(simulationOutput\_richness\_by\_type)



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



# RICHNESS ACROSS HARVESTED UNHARVESTED

## run glm for harvest on species richness using Poisson dist

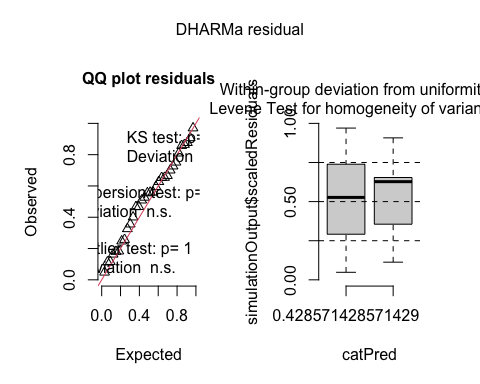
## SIG HIGHER RICHNESS IN NON\_HARVESTED SITES

higher richness in non-harvested sites - using Poisson dist - including elevation as random effect in glm

richness\_by\_harvest\_glm <- glm(no\_species ~ harvested + (1|Elevation),   
 data = richness\_by\_site,   
 family = poisson(link = "log"))  
summary(richness\_by\_harvest\_glm)

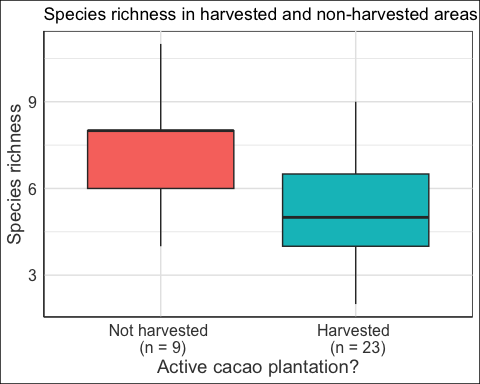
##   
## Call:  
## glm(formula = no\_species ~ harvested + (1 | Elevation), family = poisson(link = "log"),   
## data = richness\_by\_site)  
##   
## Coefficients: (1 not defined because of singularities)  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 1.9924 0.1231 16.187 <2e-16 \*\*\*  
## harvestedYes -0.3658 0.1539 -2.376 0.0175 \*   
## 1 | ElevationTRUE NA NA NA NA   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for poisson family taken to be 1)  
##   
## Null deviance: 25.719 on 31 degrees of freedom  
## Residual deviance: 20.287 on 30 degrees of freedom  
## AIC: 137.6  
##   
## Number of Fisher Scoring iterations: 4

# within group deviation from uniformity n.s.  
# Levene Test for homogeneity of variance n.s.  
simulationOutput\_richness\_by\_harvest <- simulateResiduals(  
 fittedModel = richness\_by\_harvest\_glm)  
plot(simulationOutput\_richness\_by\_harvest)



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



# RAREFIED RICHNESS ACROSS LANDUSE TYPES

need to go through rarefy\_scrap to piece this piece together - use Chao iNEXT package

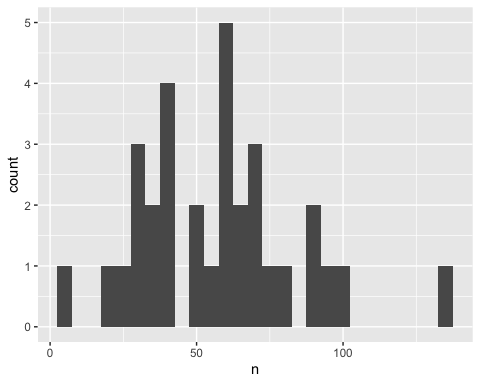
# ABUNDANCE ACROSS LANDUSE TYPES

here just simply number of new individuals - could eventually complicate this to look at pop size estimates?

## before selecting statistical test, need to look at dist. of abundance by site

* use Poisson dist. b/c using count data with no upper limit!

(abundance\_by\_site\_hist <- ggplot(abundance\_by\_site, aes(n)) +  
 geom\_histogram(binwidth = 5))



## run glm for abundance across landuse types, using Gaussian dist. Elev rand eff

DONT RUN POISSON HERE! - quantile deviations detected - combined adjusted quantile test significant

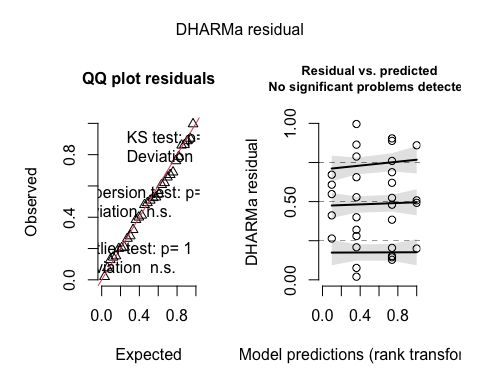
## NO SIG DIFF IN ABUNDANCE ACROSS LANDUSE TYPES

* elevation random effect

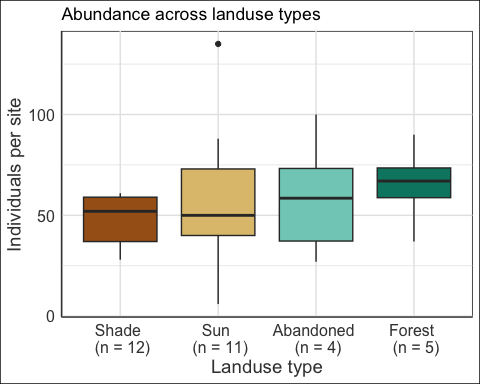
# default family dist is Gaussian  
abundance\_by\_type\_glm <- glm(n ~ Tipo + (1|Elevation), data = abundance\_by\_site)  
summary(abundance\_by\_type\_glm)

##   
## Call:  
## glm(formula = n ~ Tipo + (1 | Elevation), data = abundance\_by\_site)  
##   
## Coefficients: (1 not defined because of singularities)  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 47.400 12.260 3.866 0.000601 \*\*\*  
## TipoC 9.782 14.786 0.662 0.513659   
## TipoN 12.017 14.592 0.824 0.417181   
## TipoV 17.850 18.390 0.971 0.340034   
## 1 | ElevationTRUE NA NA NA NA   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for gaussian family taken to be 751.518)  
##   
## Null deviance: 21838 on 31 degrees of freedom  
## Residual deviance: 21043 on 28 degrees of freedom  
## AIC: 308.45  
##   
## Number of Fisher Scoring iterations: 2

# no significant problems detected w/Gaussian!  
simulationOutput\_abundance\_by\_type <- simulateResiduals(  
 fittedModel = abundance\_by\_type\_glm)  
plot(simulationOutput\_abundance\_by\_type)



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

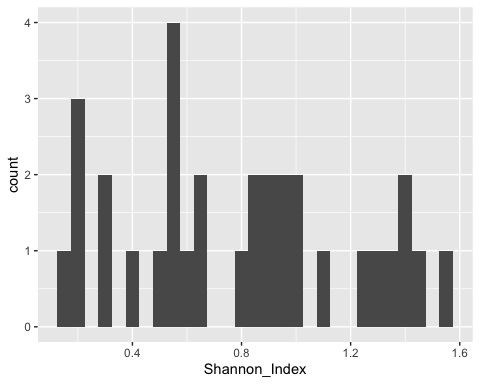


# DIVERSITY ACROSS LANDUSE TYPES

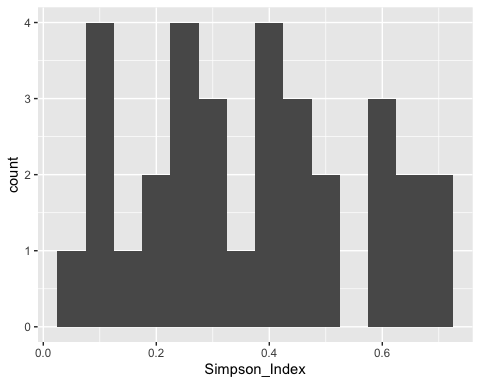
Look at how different diversity metrics compare across landuse types

## look at dist of diversity metrics

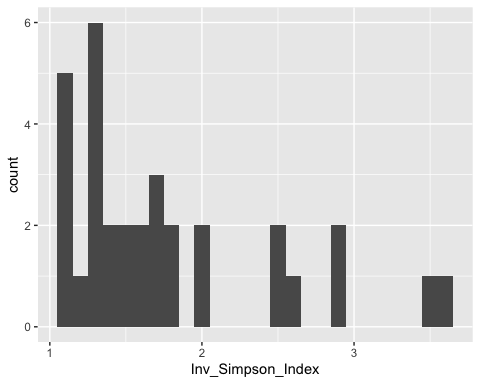
# shannon - clearly not normal dist!   
(shannon\_hist <- ggplot(diversity\_by\_site, aes(Shannon\_Index)) +  
 geom\_histogram(binwidth = .05))



# simpson - clearly not normal dist!   
(simpson\_hist <- ggplot(diversity\_by\_site, aes(Simpson\_Index)) +  
 geom\_histogram(binwidth = .05))



# inv\_simpson - clearly not normal dist!   
(inv\_simpson\_hist <- ggplot(diversity\_by\_site, aes(Inv\_Simpson\_Index)) +  
 geom\_histogram(binwidth = .1))



## use Gaussian dist for Shannon, Simpson, inv\_Gaussian for Inv\_Simpson

Thought Poisson dist would work for these b/c not normal and from count data - woah! sig deviation for Poisson dist. - quantile deviations detected - combined adjusted quantile test sig

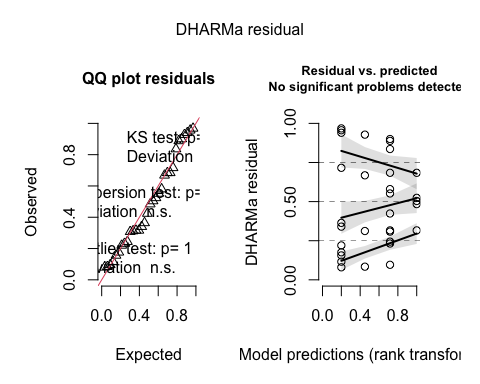
## SIG HIGHER SHANNON DIV IN BOSQUE

significantly higher shannon div in Bosque vs. other sites - viejo is close

# shannon w/ Gaussian dist - all good  
shannon\_by\_type\_glm <- glm(Shannon\_Index ~ Tipo + (1|Elevation),   
 data = diversity\_by\_site)  
summary(shannon\_by\_type\_glm)

##   
## Call:  
## glm(formula = Shannon\_Index ~ Tipo + (1 | Elevation), data = diversity\_by\_site)  
##   
## Coefficients: (1 not defined because of singularities)  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.88878 0.19028 4.671 6.82e-05 \*\*\*  
## TipoC -0.18268 0.22949 -0.796 0.433   
## TipoN -0.04856 0.22648 -0.214 0.832   
## TipoV -0.09221 0.28543 -0.323 0.749   
## 1 | ElevationTRUE NA NA NA NA   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for gaussian family taken to be 0.1810409)  
##   
## Null deviance: 5.2245 on 31 degrees of freedom  
## Residual deviance: 5.0691 on 28 degrees of freedom  
## AIC: 41.85  
##   
## Number of Fisher Scoring iterations: 2

# no significant problems detected using Gaussian dist!  
simulationOutput\_shannon\_by\_type <- simulateResiduals(  
 fittedModel = shannon\_by\_type\_glm)  
plot(simulationOutput\_shannon\_by\_type)



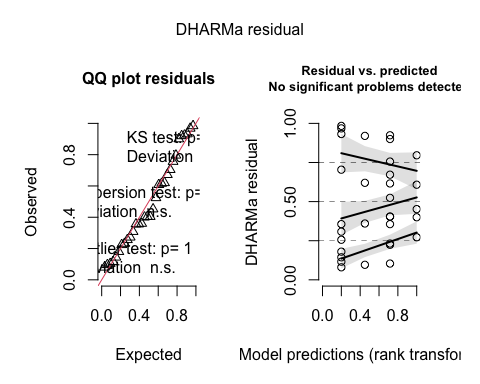
## SIG HIGHER SIMPSON DIV IN BOSQUE

sig higher simpson div in Bosque vs. other sites - viejo not particularly close

# simpson w/ Gaussian dist  
simpson\_by\_type\_glm <- glm(Simpson\_Index ~ Tipo + (1|Elevation),   
 data = diversity\_by\_site)  
summary(simpson\_by\_type\_glm)

##   
## Call:  
## glm(formula = Simpson\_Index ~ Tipo + (1 | Elevation), data = diversity\_by\_site)  
##   
## Coefficients: (1 not defined because of singularities)  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.44154 0.09068 4.869 3.97e-05 \*\*\*  
## TipoC -0.11867 0.10937 -1.085 0.287   
## TipoN -0.06771 0.10793 -0.627 0.536   
## TipoV -0.08283 0.13602 -0.609 0.547   
## 1 | ElevationTRUE NA NA NA NA   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for gaussian family taken to be 0.041116)  
##   
## Null deviance: 1.2012 on 31 degrees of freedom  
## Residual deviance: 1.1512 on 28 degrees of freedom  
## AIC: -5.5844  
##   
## Number of Fisher Scoring iterations: 2

# no significant problems detected using Gaussian dist!  
simulationOutput\_simpson\_by\_type <- simulateResiduals(  
 fittedModel = simpson\_by\_type\_glm)  
plot(simulationOutput\_simpson\_by\_type)



## NO SIG DIFF IN INV\_SIMPSON DIV ACROSS LANDUSE TYPES

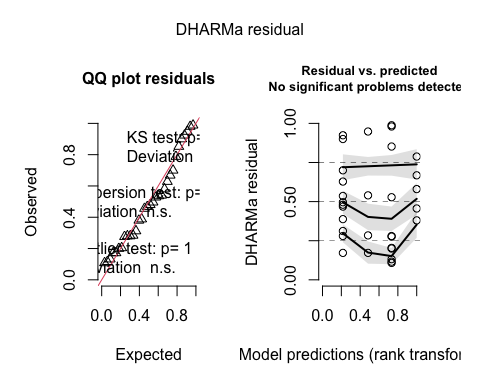
no sig difference in inv\_simpson diversity across landtypes - Bosque .18, viejo not close

NOTE THAT HERE USING INVERSE GAUSSIAN DIST - Gaussian - quantile deviations detected - comb adjust - quant test sig - Binomial - error - y values must be b/ween 0:1 - Gamma - deviation significant w/ outlier test but better than others - Inv Gaussian - works! need to load additional cran package SuppDists

# inv\_simpson w/ inverse gaussian dist.  
inv\_simpson\_by\_type\_glm <- glm(Inv\_Simpson\_Index ~ Tipo + (1|Elevation),   
 data = diversity\_by\_site,  
 family = inverse.gaussian(link = "1/mu^2"))  
summary(inv\_simpson\_by\_type\_glm)

##   
## Call:  
## glm(formula = Inv\_Simpson\_Index ~ Tipo + (1 | Elevation), family = inverse.gaussian(link = "1/mu^2"),   
## data = diversity\_by\_site)  
##   
## Coefficients: (1 not defined because of singularities)  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.29149 0.10943 2.664 0.0127 \*  
## TipoC 0.01875 0.13398 0.140 0.8897   
## TipoN 0.04811 0.13509 0.356 0.7244   
## TipoV 0.02471 0.16996 0.145 0.8854   
## 1 | ElevationTRUE NA NA NA NA   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for inverse.gaussian family taken to be 0.09511237)  
##   
## Null deviance: 2.2647 on 31 degrees of freedom  
## Residual deviance: 2.2511 on 28 degrees of freedom  
## AIC: 64.69  
##   
## Number of Fisher Scoring iterations: 6

# no significant problems detected using inverse Gaussian dist!  
# all other tested dist threw errors w/ DHARMa  
simulationOutput\_inv\_simpson\_by\_type <- simulateResiduals(  
 fittedModel = inv\_simpson\_by\_type\_glm)  
plot(simulationOutput\_inv\_simpson\_by\_type)

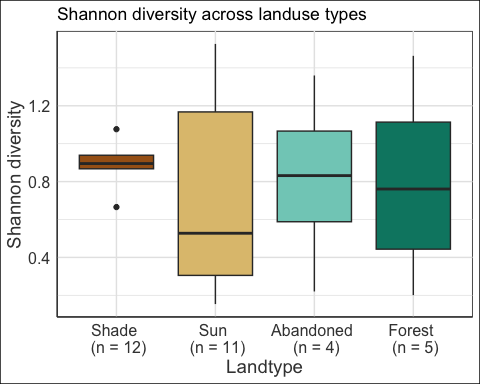


## plot how diversity indices vary across land types

* these are pretty much identical regardless of index used

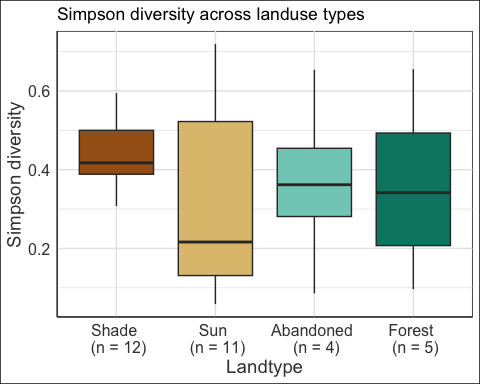
shannon

plot\_shannon <- ggplot(diversity\_by\_site,   
 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



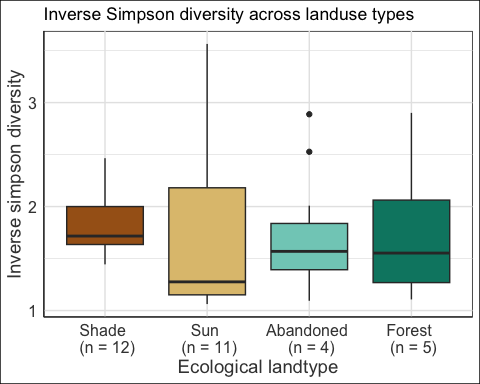
simpson

plot\_simpson <- ggplot(diversity\_by\_site,   
 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(diversity\_by\_site,   
 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



# DIVERSITY ACROSS HARVEST, UNHARVESTED

do the same thing as above, compare each index using harvested, unharvested column - glms across harvested, unharvested

## use Gaussian dist for Shannon, Simpson, Inv\_Simpson

NOTE THAT HERE THE GAUSSIAN DIST WORKS FINE FOR INV\_SIMPSON

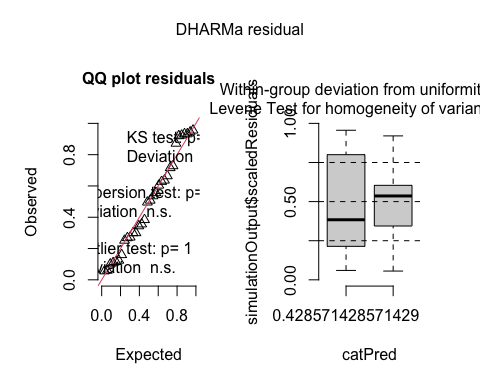
## SIG HIGHER SHANNON DIV IN NON-HARVESTED SITES

check w/ interpretation of this glm summary

# using Gaussian dist.  
shannon\_harvest\_glm <- glm(Shannon\_Index ~ harvested + (1|Elevation),   
 data = diversity\_by\_site)  
summary(shannon\_harvest\_glm)

##   
## Call:  
## glm(formula = Shannon\_Index ~ harvested + (1 | Elevation), data = diversity\_by\_site)  
##   
## Coefficients: (1 not defined because of singularities)  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.84779 0.13866 6.114 1.01e-06 \*\*\*  
## harvestedYes -0.07172 0.16356 -0.439 0.664   
## 1 | ElevationTRUE NA NA NA NA   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for gaussian family taken to be 0.1730423)  
##   
## Null deviance: 5.2245 on 31 degrees of freedom  
## Residual deviance: 5.1913 on 30 degrees of freedom  
## AIC: 38.612  
##   
## Number of Fisher Scoring iterations: 2

# no issues with Gaussian dist  
simulationOutput\_shannon\_harvest <- simulateResiduals(  
 fittedModel = shannon\_harvest\_glm)  
plot(simulationOutput\_shannon\_harvest)



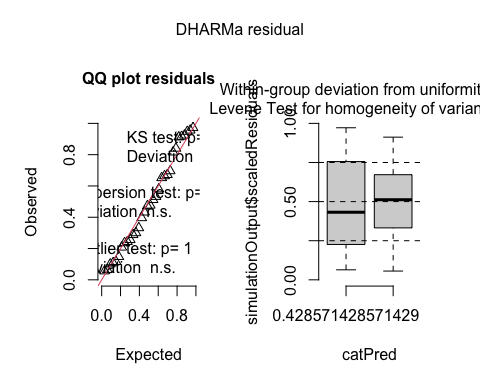
## SIG HIGHER SIMPSON DIV IN NON-HARVESTED SITES

check interpretation of this glm summary

# using Gaussian dist.  
simpson\_harvest\_glm <- glm(Simpson\_Index ~ harvested + (1|Elevation),   
 data = diversity\_by\_site)  
summary(simpson\_harvest\_glm)

##   
## Call:  
## glm(formula = Simpson\_Index ~ harvested + (1 | Elevation), data = diversity\_by\_site)  
##   
## Coefficients: (1 not defined because of singularities)  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.40473 0.06615 6.119 1e-06 \*\*\*  
## harvestedYes -0.05527 0.07802 -0.708 0.484   
## 1 | ElevationTRUE NA NA NA NA   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for gaussian family taken to be 0.03937981)  
##   
## Null deviance: 1.2012 on 31 degrees of freedom  
## Residual deviance: 1.1814 on 30 degrees of freedom  
## AIC: -8.7572  
##   
## Number of Fisher Scoring iterations: 2

# no issues with Gaussian dist  
simulationOutput\_simpson\_harvest <- simulateResiduals(  
 fittedModel = simpson\_harvest\_glm)  
plot(simulationOutput\_simpson\_harvest)



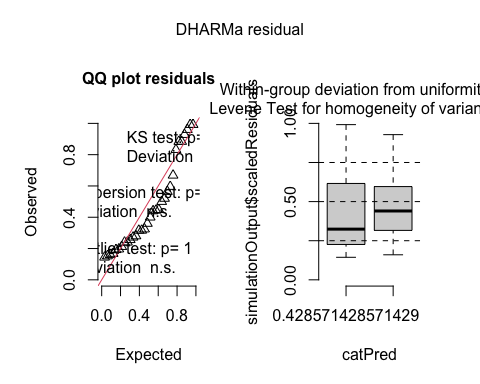
## NO SIG DIFF IN IN\_SIMPSON ACROSS HARVEST, NON-HARVEST

check interpretation of this glm summary

# using Gaussian dist.  
inv\_simpson\_harvest\_glm <- glm(Inv\_Simpson\_Index ~ harvested + (1|Elevation),   
 data = diversity\_by\_site)  
summary(inv\_simpson\_harvest\_glm)

##   
## Call:  
## glm(formula = Inv\_Simpson\_Index ~ harvested + (1 | Elevation),   
## data = diversity\_by\_site)  
##   
## Coefficients: (1 not defined because of singularities)  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.81937 0.23584 7.715 1.32e-08 \*\*\*  
## harvestedYes -0.06544 0.27818 -0.235 0.816   
## 1 | ElevationTRUE NA NA NA NA   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for gaussian family taken to be 0.5005671)  
##   
## Null deviance: 15.045 on 31 degrees of freedom  
## Residual deviance: 15.017 on 30 degrees of freedom  
## AIC: 72.602  
##   
## Number of Fisher Scoring iterations: 2

# no issues with Gaussian dist  
simulationOutput\_inv\_simpson\_harvest <- simulateResiduals(  
 fittedModel = inv\_simpson\_harvest\_glm)  
plot(simulationOutput\_inv\_simpson\_harvest)

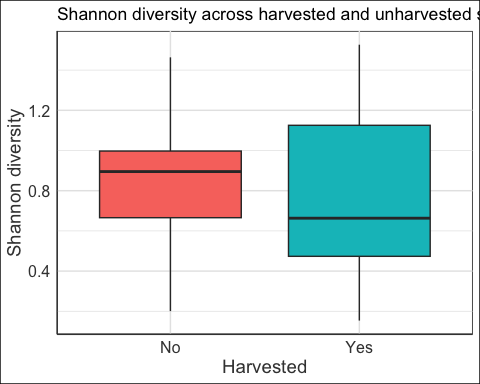


## make figures for each

* aesthetically think about fill colors

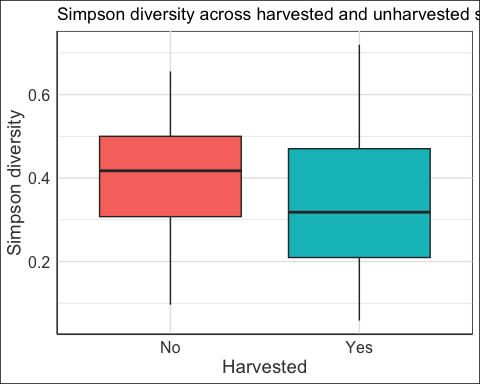
shannon\_harvest

plot\_shannon\_harvest <- ggplot(diversity\_by\_site,   
 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



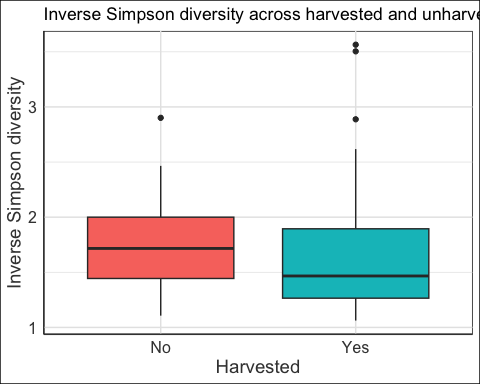
simpson\_harvest

plot\_simpson\_harvest <- ggplot(diversity\_by\_site,   
 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(diversity\_by\_site,   
 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



# COMPOSITION ACROSS LANDUSE TYPES

Variety of ways to look at this, can’t do all of them! - NMDS - more for visualizing differences - start here - Bray-Curtiss Dissimilarity Index - quantifies dissimilarity - 0:1 (same:diff) - this one preferred b/c incorporates species abundances - Jaccard Similarity Index - doesn’t incorporate species abundances - Analysis of similarities for statistics

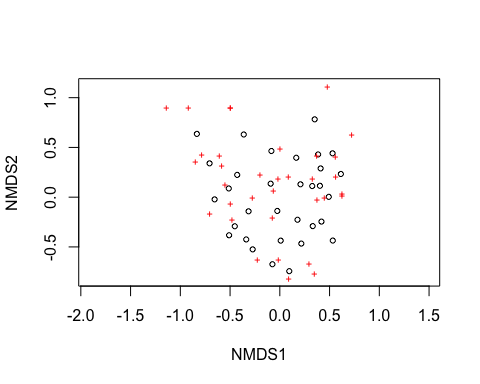
## create nmds plot and draw boundaries of landuse types

## WHY ARE NUMBERS INSTEAD OF SITES SHOWING?

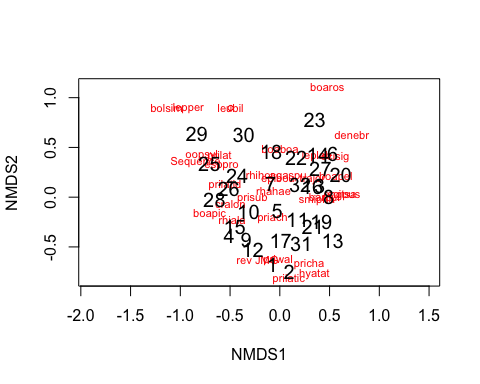
# assign treatment levels per landuse type on reordered new\_mat\_reordered  
landuse\_treatments=c(rep("Shade",12),  
 rep("Sun",11),   
 rep("Abandoned",4),   
 rep("Forest",5))  
# ensure that labels are in correct order and have correct length  
  
# use community by species matrix  
new\_mat\_NMDS = metaMDS(new\_mat, k = 2)

## Square root transformation  
## Wisconsin double standardization  
## Run 0 stress 0.2494371   
## Run 1 stress 0.2613206   
## Run 2 stress 0.2692783   
## Run 3 stress 0.2610522   
## Run 4 stress 0.2925289   
## Run 5 stress 0.2494871   
## ... Procrustes: rmse 0.01166604 max resid 0.04697205   
## Run 6 stress 0.2650125   
## Run 7 stress 0.2510953   
## Run 8 stress 0.2962829   
## Run 9 stress 0.2494873   
## ... Procrustes: rmse 0.01160095 max resid 0.04667685   
## Run 10 stress 0.2611437   
## Run 11 stress 0.249437   
## ... New best solution  
## ... Procrustes: rmse 6.275162e-05 max resid 0.0002482263   
## ... Similar to previous best  
## Run 12 stress 0.2610522   
## Run 13 stress 0.2494873   
## ... Procrustes: rmse 0.01175921 max resid 0.04743315   
## Run 14 stress 0.2494871   
## ... Procrustes: rmse 0.01162793 max resid 0.04683503   
## Run 15 stress 0.2609532   
## Run 16 stress 0.2494871   
## ... Procrustes: rmse 0.01163024 max resid 0.04684584   
## Run 17 stress 0.2494371   
## ... Procrustes: rmse 0.0001442166 max resid 0.0005849329   
## ... Similar to previous best  
## Run 18 stress 0.2494872   
## ... Procrustes: rmse 0.01162829 max resid 0.04684118   
## Run 19 stress 0.3008624   
## Run 20 stress 0.2576728   
## \*\*\* Best solution repeated 2 times

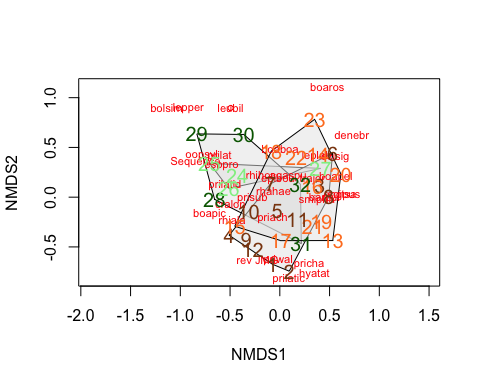
# sites are open circles, species are red crosses  
plot(new\_mat\_NMDS)



# 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)



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)

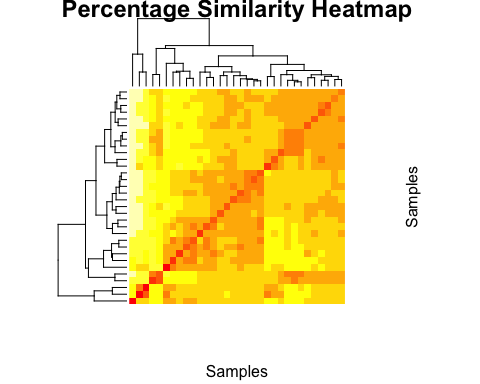


## calculate Bray-Curtis Index

* maybe move this to ranas\_process actually!

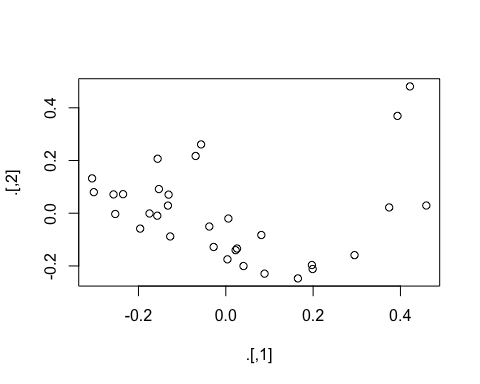
## WHY ARENT ROW NAMES BEING USED? HEATMAP UNINTELLIGBLE

# Step 1: Calculate Bray-Curtis distance  
bray\_matrix <- vegdist(new\_mat, method="bray")  
  
# Step 2: Convert distance matrix to percentage similarity  
# Convert distances to similarities (1 - distance)  
similarity\_matrix <- 1 - as.matrix(bray\_matrix)  
  
# Step 3: Plot the percentage similarity matrix  
# Example heatmap plot  
heatmap(similarity\_matrix,   
 main="Percentage Similarity Heatmap",  
 xlab="Samples", ylab="Samples",  
 labRow=NA, labCol=NA,  
 col=rev(heat.colors(10))) # Reversed heat colors for better visualization



## adjacent, random piece from Riffomonas project

# ordination of bray\_matrix using cmdscale  
  
cmdscale(bray\_matrix) %>% plot()



pcoa <- cmdscale(bray\_matrix)  
colnames(pcoa) <- c("pcoa1", "pcoa2")  
pcoa %>% head()

## pcoa1 pcoa2  
## [1,] 0.088726230 -0.22921683  
## [2,] 0.165138116 -0.24728275  
## [3,] -0.056576265 0.26111133  
## [4,] -0.302220970 0.07998492  
## [5,] 0.003766672 -0.17497837  
## [6,] 0.393206351 0.36925944