

Montana Statewide nifH Analysis

Alexander Alleman

January 23, 2019

```
## [1] "2019-02-01-12-20-31"
```

Working Directory

```
## [1] "/home/peterslab/Alex Alleman/Statewide Microbiome Analysis/Statewide analysis"
```

```
set.seed(8765)
```

Load packages

```
library(ggplot2)
library(data.table)
library(vegan)
library(ggvegan)
library(dplyr)
library(scales)
library(grid)
library(reshape2)
library(phyloseq)
library(ggpubr)
library(RColorBrewer)
library(ape)
library(grid)
library(knitr)
library(igraph)
library(Matrix)
library(ggnetwork)
library(intergraph)
library(Hmisc)
library(parallel)
library(ggrepel)
library(tinytex)
```

Colors

```
farm_col<-(c("#8c510a", "#d8b365", "#f6e8c3", "#f5f5f5", "#c7eae5", "#5ab4ac", "#01665e"))
farm_col_dark<-brewer.pal(7, "Dark2")
farm_col_paired<-brewer.pal(10, "Paired")
```

Load OTU, Taxa, and Meta data

Add OTU table with sample names on top and OTU names as row names

```
OTU_nifH<- read.delim(
  "~/Alex Alleman/Statewide Microbiome Analysis/Statewide analysis/nifH_OTUallspring.csv",
  row.names = 1)
head(OTU_nifH)[,1:10]
```

	JZ017	JZ018	JZ019	JZ020	JZ021	JZ022	JZ023	JZ024	JZ025	JZ026
OTU1	21	32923	48199	22	117	103	147	4054	5597	246
OTU2	35281	5	15	1075	101	1335	26249	376	1705	256
OTU3	11	14	13	49	20	48	9	8	16	9
OTU4	80	4	14	2576	144	136	150	129	29299	77807
OTU5	6	2	7	146	749	526	432	1115	13256	0
OTU6	19317	11274	799	748	1532	348	32	21	28	2517

These taxa were also created by Mr. DNA through a blast program and the NCBI database

```
tax_nifH<- read.delim(
  "~/Alex Alleman/Statewide Microbiome Analysis/Statewide analysis/nifH_OTU_ids_2016_fixed1.txt",
  row.names = 1)
head(tax_nifH)[,1:8]
```

	kingdom	phylum	class	order	family	genus
OTU1	k__bacteria	p__firmicutes	c__bacilli	o__bacillales	f__paenibacillaceae	g__pae
OTU2	k__bacteria	p__firmicutes	c__bacilli	o__bacillales	f__paenibacillaceae	g__pae
OTU3	k__bacteria	p__proteobacteria	c__alphaproteobacteria	o__rhizobiales	f__rhizobiaceae	g__rhiz
OTU4	k__bacteria	p__firmicutes	c__bacilli	o__bacillales	f__paenibacillaceae	g__pae
OTU5	k__bacteria	p__proteobacteria	c__alphaproteobacteria	o__rhizobiales	f__bradyrhizobiaceae	g__bra
OTU6	k__bacteria	p__proteobacteria	c__alphaproteobacteria	o__rhizobiales	f__bradyrhizobiaceae	g__rho

Meta data set has be placed together from all the spring and summer data with excel

```
meta<- read.delim(
  "~/Alex Alleman/Statewide Microbiome Analysis/Statewide analysis/all_metadata_summer.csv",
  colClasses = c(rep('factor', 7), rep('numeric', 3), rep('factor', 4), 'numeric',
    rep('factor', 3), rep('numeric', 28) ) )
head(meta)[,1:5]
```

	Site	ARC	Season	Sample_dates	Pea_variety
JZ032	Kalispell	NWARC	Summer	2016-summer	Delta
JZ031	Kalispell	NWARC	Summer	2016-summer	CDC Saffron
JZ030	Kalispell	NWARC	Summer	2016-summer	AC Earlystar
JZ034	Kalispell	NWARC	Summer	2016-summer	Majoret
JZ033	Kalispell	NWARC	Summer	2016-summer	DS Admiral
JZ035	Kalispell	NWARC	Summer	2016-summer	Navarro

Removed all Havre for analysis because samples were taken in 6" sections instead of 12"

```
meta2 <- meta[-c(19:48),]  
sapply(meta2, class)
```

```
##           Site           ARC           Season   Sample_dates  
##           "factor"        "factor"        "factor"        "factor"  
##   Pea_variety           Plot   season_precip   irrigation  
##           "factor"        "factor"        "numeric"        "numeric"  
## total_precip_irr   sample_depth           Date   Tillage  
##           "numeric"        "factor"        "factor"        "factor"  
##   prev_crop       grain_yield   elevation           lat  
##           "factor"        "numeric"        "factor"        "factor"  
##           lon   Organic_Matter   Moisture_Content   Nitrate_Nitrite  
##           "factor"        "numeric"        "numeric"        "numeric"  
##   Ammonia   Av_Phosphorus   Av_Potassium   Sulfate_Sulfur  
##           "numeric"        "numeric"        "numeric"        "numeric"  
##           pH           Boron           Arsenic           Barium  
##           "numeric"        "numeric"        "numeric"        "numeric"  
##   Cadmium           Calcium           Chromium           Cobalt  
##           "numeric"        "numeric"        "numeric"        "numeric"  
##   Copper           Iron           Lead           Magnesium  
##           "numeric"        "numeric"        "numeric"        "numeric"  
##   Manganese   Molybdenum           Nickel           Phosphorus  
##           "numeric"        "numeric"        "numeric"        "numeric"  
##   Potassium           Sodium           Sulfur           Vanadium  
##           "numeric"        "numeric"        "numeric"        "numeric"  
##           Zinc  
##           "numeric"
```

Convert to matrix

```
OTU_nifH_m<-as.matrix(OTU_nifH)  
tax_nifH_m<-as.matrix(tax_nifH)  
meta_m<-as.matrix(meta2)  
  
class(OTU_nifH_m)
```

```
## [1] "matrix"
```

```
class(tax_nifH_m)
```

```
## [1] "matrix"
```

```
class(meta_m)
```

```
## [1] "matrix"
```

Make phyloseq object

```
OTUnifH = otu_table(OTU_nifH_m, taxa_are_rows = TRUE)
TAXnifH = tax_table(tax_nifH_m)

physeq_nifH = phyloseq(OTUnifH, TAXnifH)
```

Get physeq info

```
physeq_nifH
```

```
## phyloseq-class experiment-level object
## otu_table()   OTU Table:             [ 8821 taxa and 101 samples ]
## tax_table()   Taxonomy Table:        [ 8821 taxa by 8 taxonomic ranks ]
```

Add meta data to phyloseq object

```
meta_phy <- sample_data(meta2)
sample_names(meta_phy)
```

```
## [1] "JZ032" "JZ031" "JZ030" "JZ034" "JZ033" "JZ035" "JZ040" "JZ046"
## [9] "JZ042" "JZ044" "JZ038" "JZ036" "JZ041" "JZ047" "JZ043" "JZ045"
## [17] "JZ039" "JZ037" "JZ081" "JZ078" "JZ082" "JZ079" "JZ080" "JZ083"
## [25] "JZ105" "JZ107" "JZ103" "JZ102" "JZ106" "JZ104" "JZ084" "JZ085"
## [33] "JZ086" "JZ087" "JZ088" "JZ089" "JZ090" "JZ091" "JZ092" "JZ093"
## [41] "JZ094" "JZ095" "JZ096" "JZ097" "JZ098" "JZ099" "JZ100" "JZ101"
## [49] "JZ112" "JZ109" "JZ110" "JZ111" "JZ108" "JZ113"
```

```
physeq_nifH<-merge_phyloseq(physeq_nifH, meta_phy)
physeq_nifH
```

```
## phyloseq-class experiment-level object
## otu_table()   OTU Table:             [ 8821 taxa and 54 samples ]
## sample_data() Sample Data:           [ 54 samples by 45 sample variables ]
## tax_table()   Taxonomy Table:        [ 8821 taxa by 8 taxonomic ranks ]
```

```
physeq_nifH<-rarefy_even_depth(physeq_nifH)
```

```
## You set `rngseed` to FALSE. Make sure you've set & recorded
## the random seed of your session for reproducibility.
## See `?set.seed`
```

```
## ...
```

```
## 36100TUs were removed because they are no longer
## present in any sample after random subsampling
```

```
## ...
```

```
physeq_nifH
```

```
## phyloseq-class experiment-level object
## otu_table() OTU Table:      [ 5211 taxa and 54 samples ]
## sample_data() Sample Data:  [ 54 samples by 45 sample variables ]
## tax_table() Taxonomy Table:  [ 5211 taxa by 8 taxonomic ranks ]
```

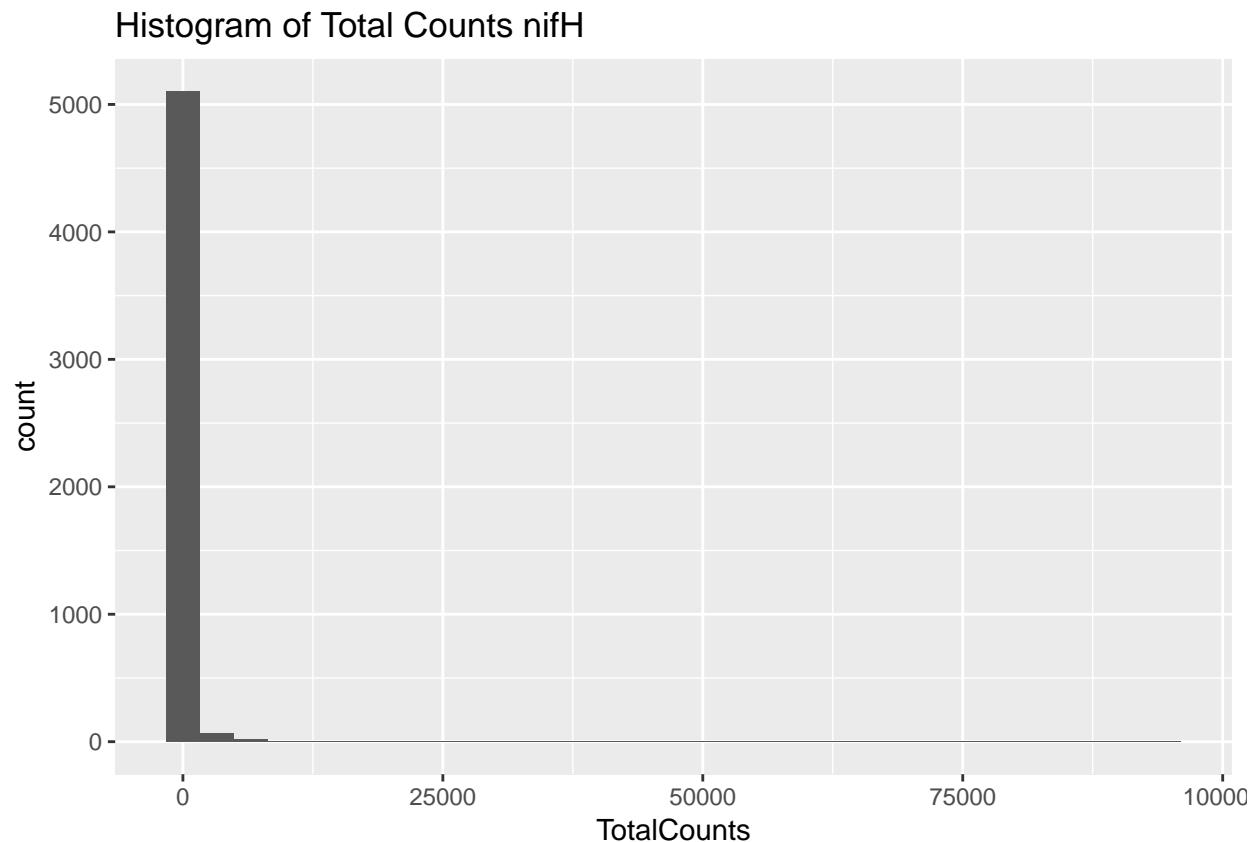
Trim data

Trim data to exclude OTUs that are not in any samples

Source of trim protocol <http://evomics.org/wp-content/uploads/2016/01/phyloseq-Lab-01-Answers.html#taxa-total-counts-histogram>

```
tdt_nifH = data.table(tax_table(physeq_nifH),
                      TotalCounts = taxa_sums(physeq_nifH),
                      OTU = taxa_names(physeq_nifH))
ggplot(tdt_nifH, aes(TotalCounts)) +
  geom_histogram() +
  ggtitle("Histogram of Total Counts nifH")
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



```
# How many OTUs have low count (Rare)?
tdt_nifH[(TotalCounts <= 0), .N]#zero count
```

```
## [1] 0
```

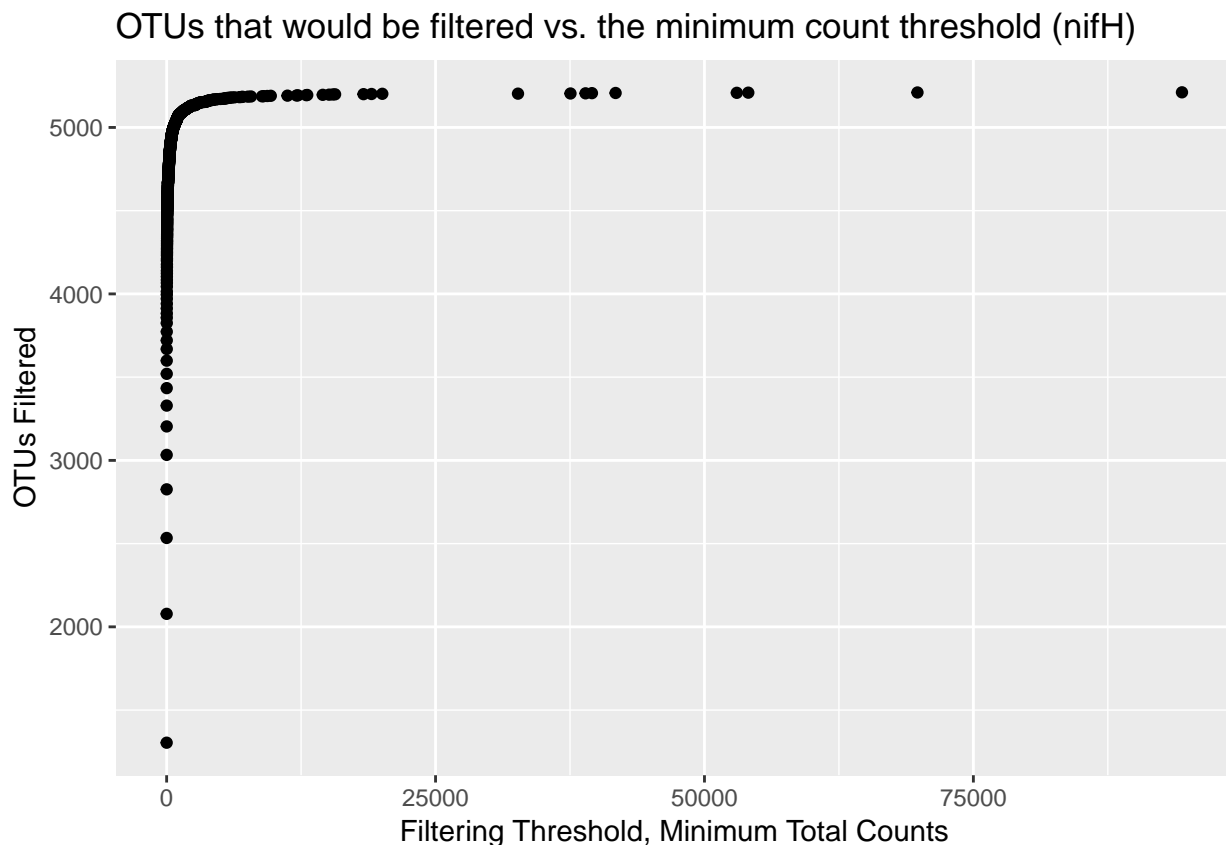
```
tdt_nifH[(TotalCounts <= 1), .N]#single count
```

```
## [1] 1304
```

```
tdt_nifH[(TotalCounts <= 2), .N]#double count
```

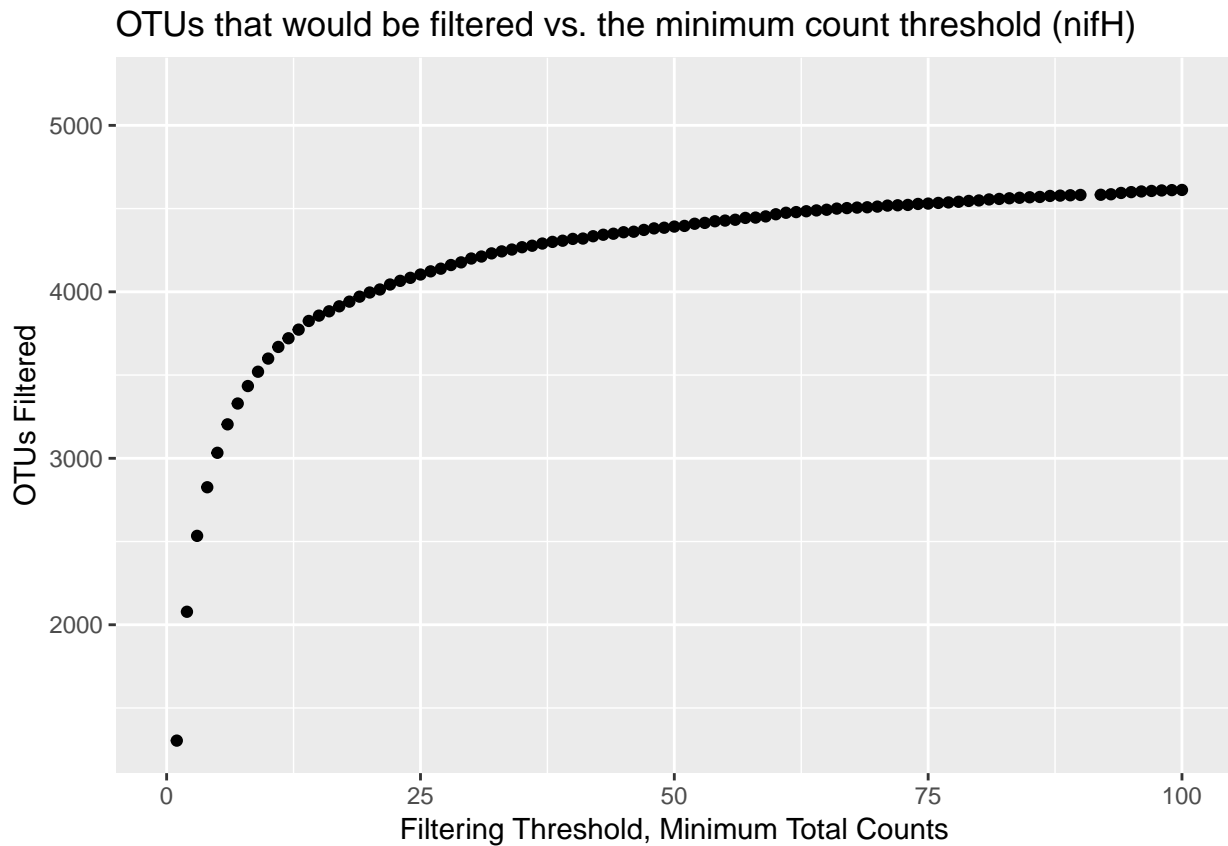
```
## [1] 2078
```

```
# taxa cumulative sum
taxcumsum_nifH = tdt_nifH[, .N, by = TotalCounts]
setkey(taxcumsum_nifH, TotalCounts)
taxcumsum_nifH[, CumSum := cumsum(N)]
# Define the plot
pCumSum_nifH = ggplot(taxcumsum_nifH, aes(TotalCounts, CumSum)) +
  geom_point() +
  xlab("Filtering Threshold, Minimum Total Counts") +
  ylab("OTUs Filtered") +
  ggtitle("OTUs that would be filtered vs. the minimum count threshold (nifH)")
pCumSum_nifH
```



Zoom in

```
pCumSum_nifH + xlim(0, 100)
```



So if we filter ~9 Total count so remove every OTU with less than 9 counts we would remove 5000 OTUs. If we remove OTUS with less than 3 counts per OTU we only remove ~2900 OTUS.

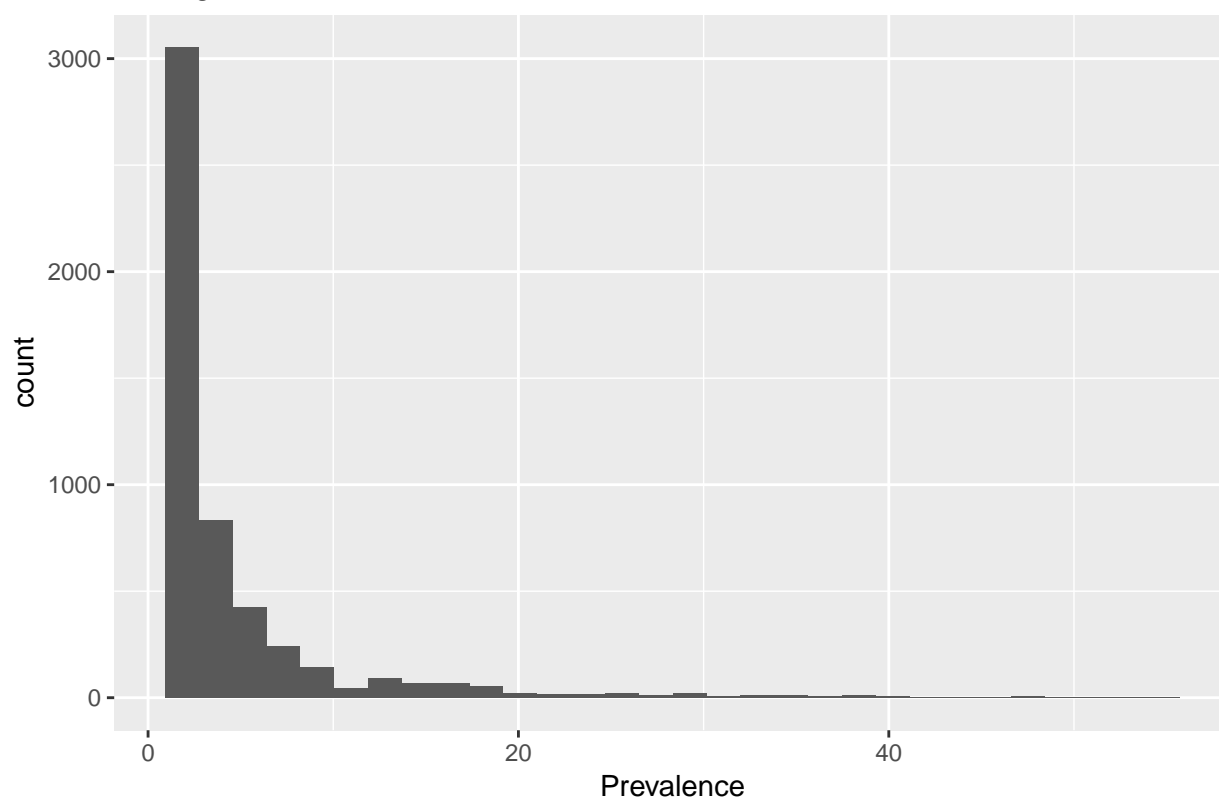
We will now to the same with OTU prevlenace (How many samples is the OTU in)

```
mdt_nifH = fast_melt(physeq_nifH)
prevdtnifH = mdt_nifH[, list(Prevalence = sum(count > 0),
                             TotalCounts = sum(count)),
                      by = TaxaID]
```

```
ggplot(prevdtnifH, aes(Prevalence)) +
  geom_histogram() +
  ggtitle("Histogram of Taxa Prevalence nifH")
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

Histogram of Taxa Prevalence nifH



```
# How many OTUS have low prevelance (Rare)?
prevdtnifH[(Prevalence <= 0), .N] #zero
```

```
## [1] 0
```

```
prevdtnifH[(Prevalence <= 1), .N] #single
```

```
## [1] 2144
```

```
prevdtnifH[(Prevalence <= 2), .N] #double
```

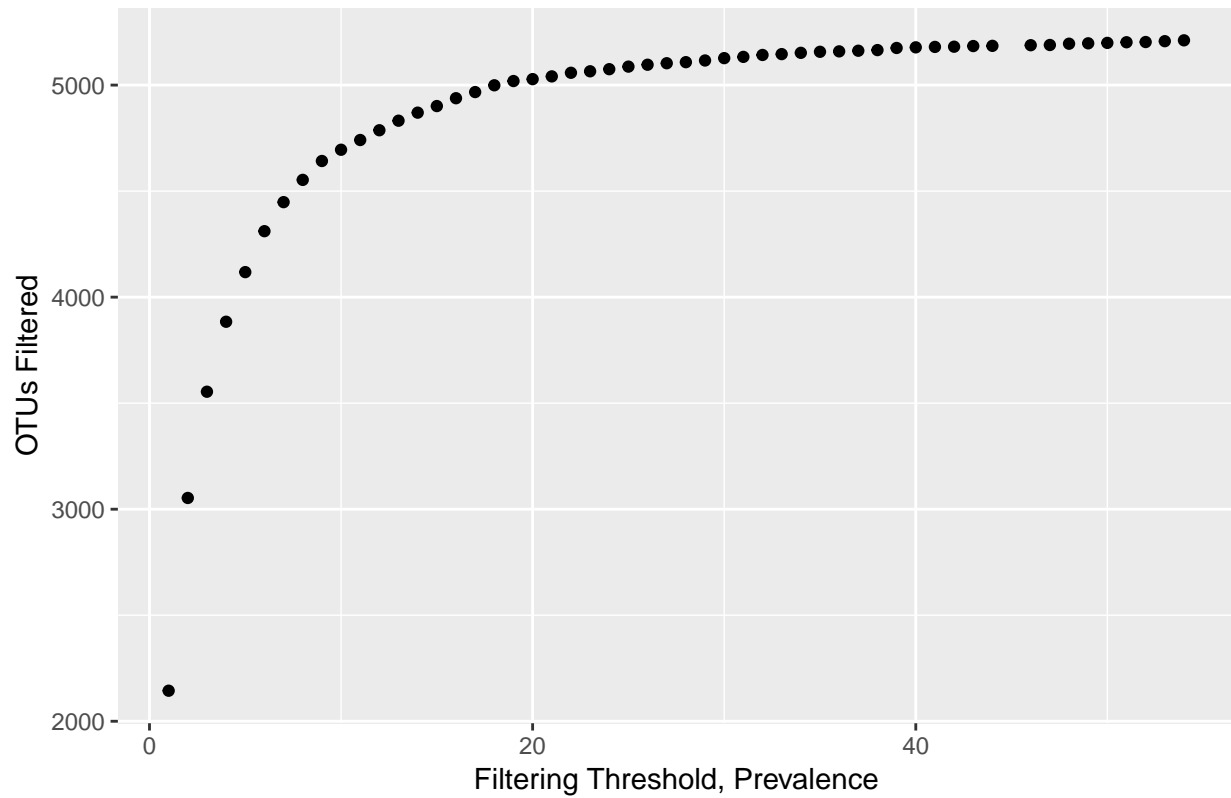
```
## [1] 3053
```

```
prevdtnifH[(Prevalence >= 54), .N] #how many OTUS are in every sample
```

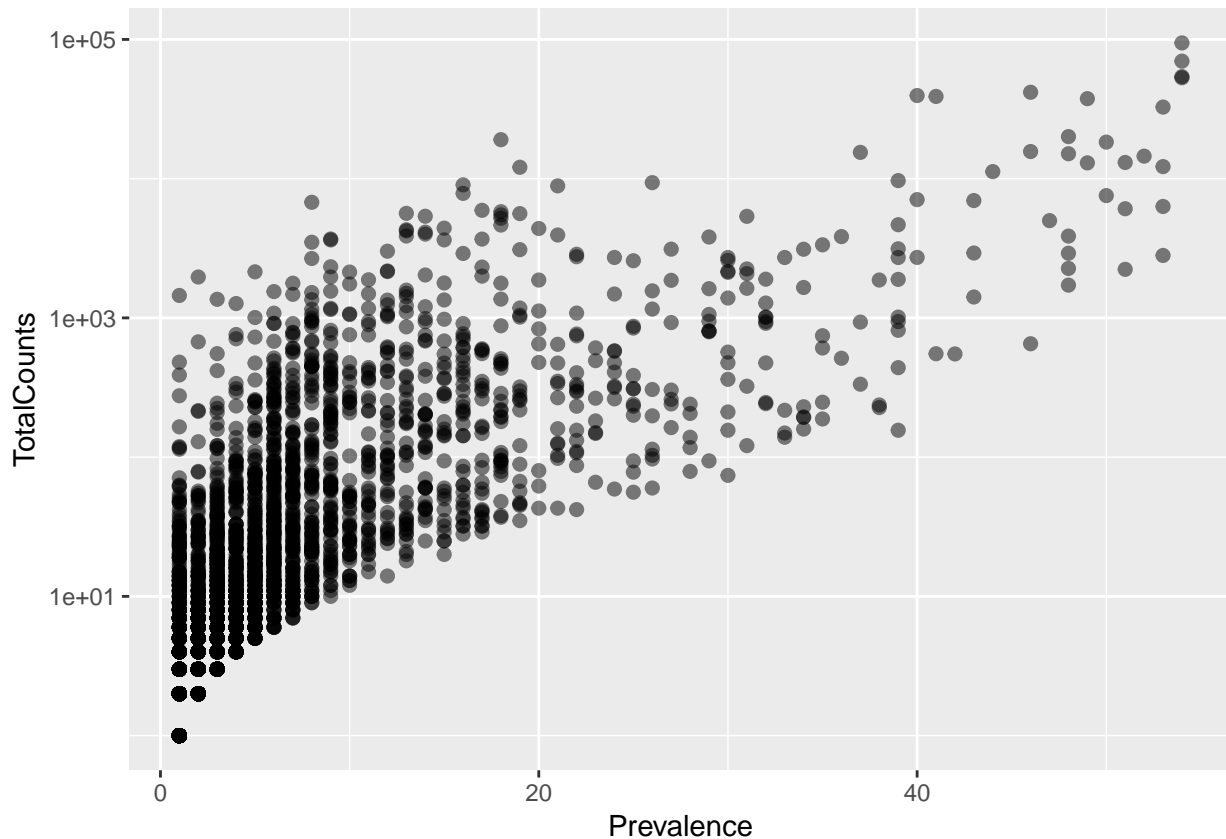
```
## [1] 4
```

```
prevcumsumnifH = prevdtnifH[, .N, by = Prevalence]
setkey(prevcumsumnifH, Prevalence)
prevcumsumnifH[, CumSum := cumsum(N)]
pPrevCumSumnifH = ggplot(prevcumsumnifH, aes(Prevalence, CumSum)) +
  geom_point() +
  xlab("Filtering Threshold, Prevalence") +
  ylab("OTUs Filtered") +
  ggtitle("OTUs that would be filtered vs. the minimum count threshold")
pPrevCumSumnifH
```


OTUs that would be filtered vs. the minimum count threshold



```
ggplot(prevdtnifH, aes(Prevalence, TotalCounts)) +  
  geom_point(size = 2, alpha = 0.5) +  
  scale_y_log10()
```



There is a good a good distribution of the OTUs the OTUs that occur in low abundnace also occur in low prevlance. By removing some of the low counts (below) we will not be losing on the similarities between sites.

Trimming

Remove less than doublets in data and prevlant in 5% of the sample

First transform to realtive abundance

```
physeq_nifH_trim = filter_taxa(physeq_nifH, function(x) sum(x > 2) > (0.05*length(x)),
                                TRUE)
physeq_nifH_trim
```

```
## phyloseq-class experiment-level object
## otu_table() OTU Table:      [ 1006 taxa and 54 samples ]
## sample_data() Sample Data:  [ 54 samples by 45 sample variables ]
## tax_table()  Taxonomy Table: [ 1006 taxa by 8 taxonomic ranks ]
```

To simplify ordination and save time we will trim the OTUs more

Remove OTUs that do not show appear more than 5 times in more than 10th of the samples

```
wh0_n = genefilter_sample(physeq_nifH_trim, filterfun_sample(function(x) x > 3),
                          A=0.1*nsamples(physeq_nifH_trim))
physeq_nifH_ord = prune_taxa(wh0_n, physeq_nifH_trim)
physeq_nifH_ord
```

```
## phyloseq-class experiment-level object
## otu_table() OTU Table: [ 446 taxa and 54 samples ]
## sample_data() Sample Data: [ 54 samples by 45 sample variables ]
## tax_table() Taxonomy Table: [ 446 taxa by 8 taxonomic ranks ]
```

Transform to even sampling depth

```
physeq_nifH_ord = transform_sample_counts(physeq_nifH_ord, function(x) 1E6 * x/sum(x))
physeq_nifH_ord
```

```
## phyloseq-class experiment-level object
## otu_table() OTU Table: [ 446 taxa and 54 samples ]
## sample_data() Sample Data: [ 54 samples by 45 sample variables ]
## tax_table() Taxonomy Table: [ 446 taxa by 8 taxonomic ranks ]
```

We have removed the majority of the low abundance data with a remaining 804 taxa which make the data analysis much more manageable.

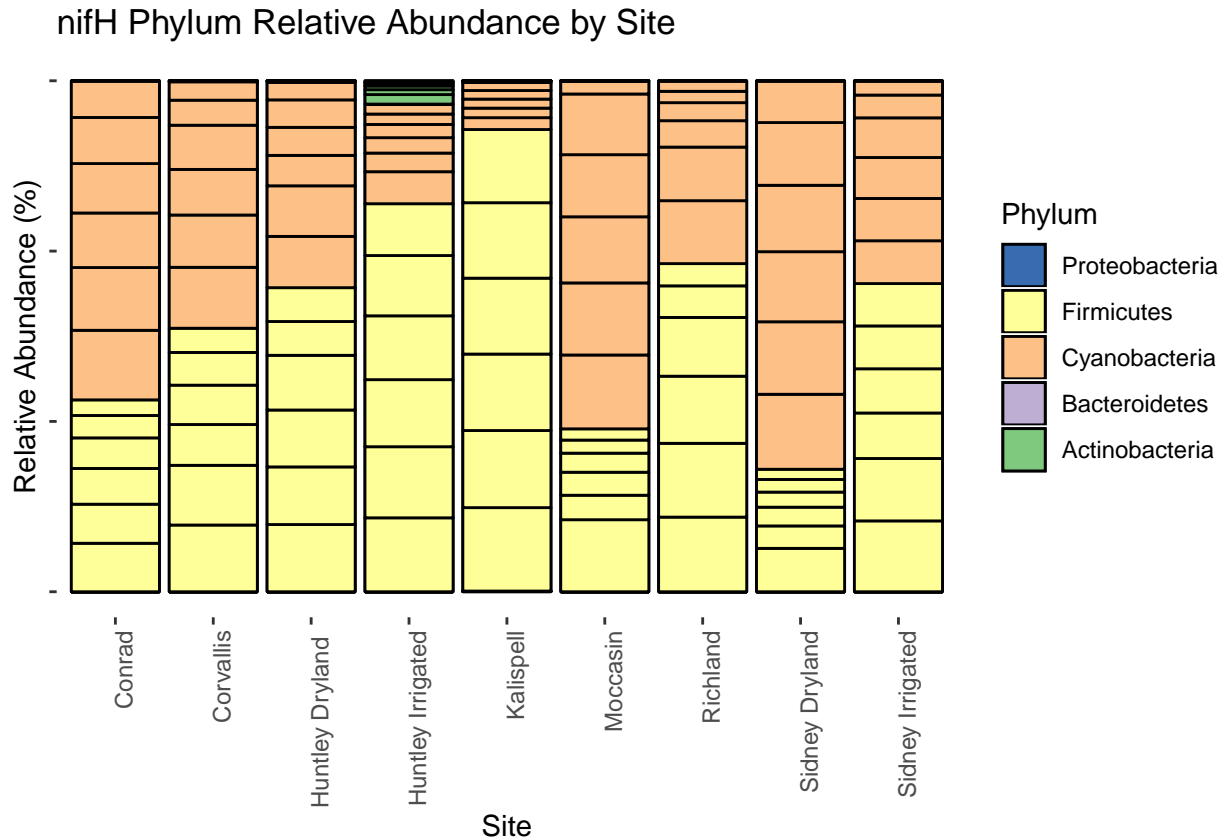
Analysis

Alpha Diveristy Analysis

Bar plots

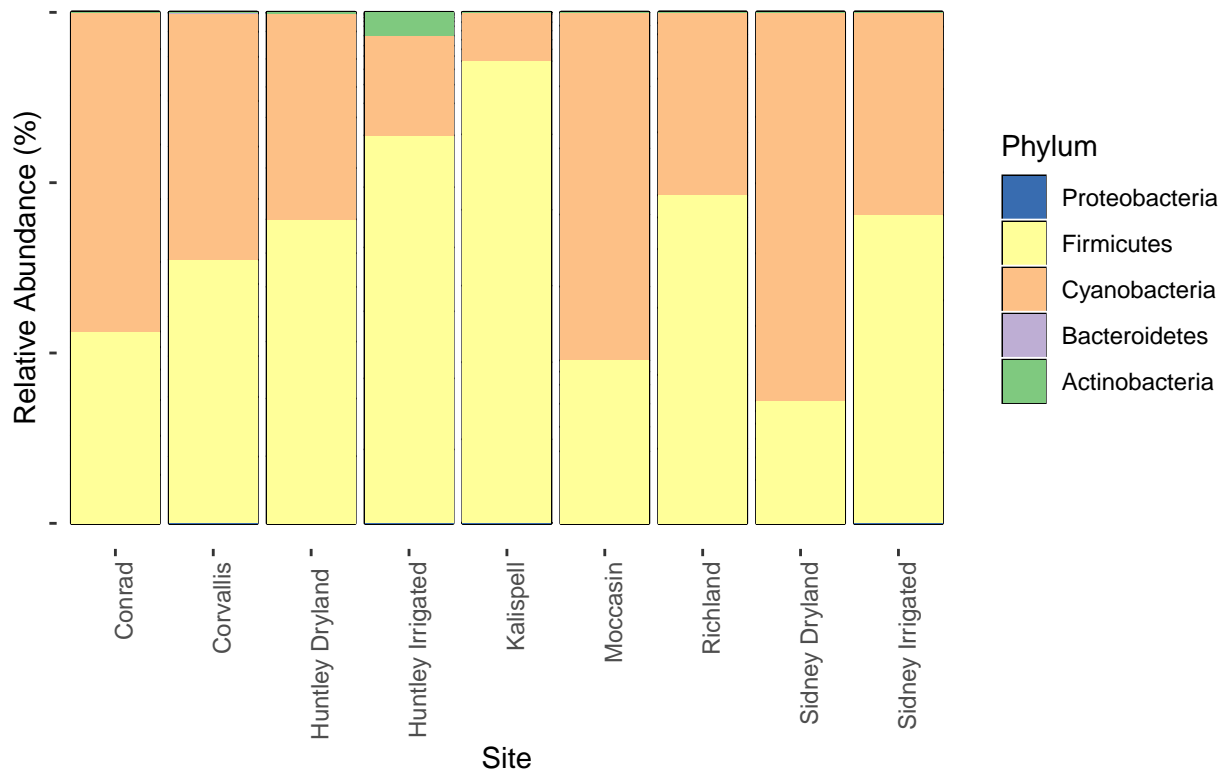
```
physeq_nifH_ord_2_phylum <- tax_glom(physeq_nifH_ord, "phylum")
plot_bar(physeq_nifH_ord_2_phylum, x = "Site", fill = "phylum")+
geom_bar(aes(fill = phylum), color = "black", stat = "identity", position = "stack",
         show.legend = TRUE)+
scale_fill_manual(name = "Phylum",
                  values=c("#7fc97f", "#beaed4", "#fdc086", "#ffff99", "#386cb0",
                           "#f0027f", "#e31a1c"),
                  labels = c("Actinobacteria", "Bacteroidetes", "Cyanobacteria",
                              "Firmicutes", "Proteobacteria", "Verrucomicrobia"),
                  guide = guide_legend(reverse = TRUE)
)+
```

```
ggtitle("nifH Phylum Relative Abundance by Site")+
ylab("Relative Abundance (%)")+
theme(axis.text.x = element_text(angle = 90, hjust = 1), axis.text.y = element_blank(),
      panel.background = element_blank())
```



```
physeq_nifH_ord_2_phylum <- tax_glom(physeq_nifH_ord, "phylum")
plot_bar(physeq_nifH_ord_2_phylum, x = "Site", fill = "phylum")+
geom_bar(aes(fill = phylum), stat = "identity", position = "stack", show.legend = TRUE)+
scale_fill_manual(name = "Phylum",
                  values=c("#7fc97f", "#beaed4", "#fdc086", "#ffff99",
                           "#386cb0", "#f0027f"),
                  labels = c("Actinobacteria", "Bacteroidetes", "Cyanobacteria",
                              "Firmicutes", "Proteobacteria", "Verrucomicrobia"),
                  guide = guide_legend(reverse = TRUE))
ggtitle("nifH Phylum Relative Abundance by Site")+
ylab("Relative Abundance (%)")+
theme(axis.text.x = element_text(angle = 90, hjust = 1), axis.text.y = element_blank(),
      panel.background = element_blank())
```

nifH Phylum Relative Abundance by Site



Reduce phylums that contribute the less than 5% of total abundance into one group

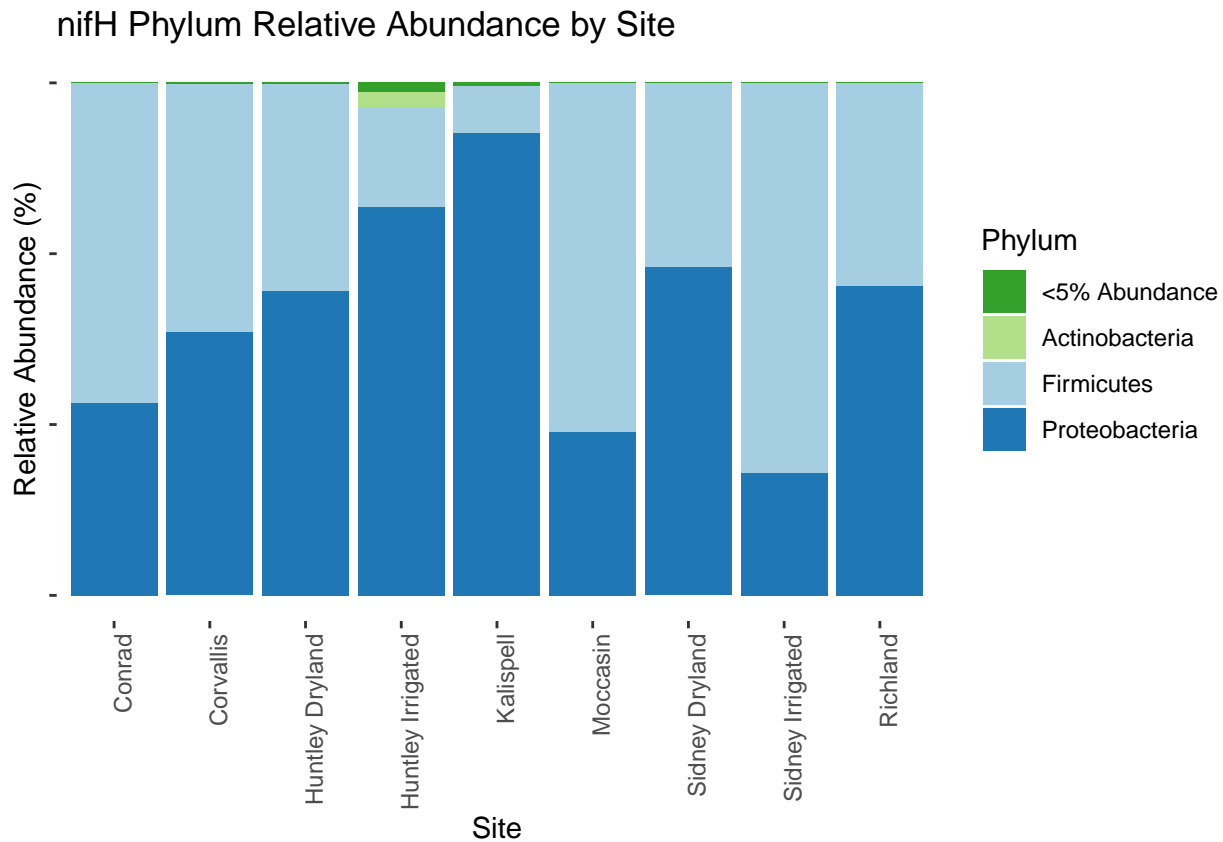
```
physeq_nifH_ord_1 = transform_sample_counts(physeq_nifH_ord, function(x) x / sum(x) )
physeq_nifH_ord_phylum <- tax_glom(physeq_nifH_ord_1, "phylum")
data_nifH_phylum <- psmelt(physeq_nifH_ord_phylum)
data_nifH_phylum$phylum <- as.character(data_nifH_phylum$phylum)
data_nifH_phylum$phylum[data_nifH_phylum$Abundance < 0.05] <- "<5% abund"

#list new phylums
unique(data_nifH_phylum$phylum)
```

```
## [1] "p__proteobacteria" "p__firmicutes"      "p__actinobacteria"
## [4] "<5% abund"
```

```
ggplot(data = data_nifH_phylum, aes(x = Site, y = Abundance, fill = phylum)) +
  geom_bar(aes(fill = phylum), stat = "identity", position = "stack", show.legend = TRUE) +
  scale_fill_manual(name = "Phylum",
    values = c("#33a02c", "#b2df8a", "#a6cee3", "#1f78b4"),
    breaks = c("p__proteobacteria", "p__firmicutes",
      "p__actinobacteria", "<5% abund" ),
    labels = c("Proteobacteria", "Firmicutes",
      "Actinobacteria", "<5% Abundance"),
    guide = guide_legend(reverse = TRUE)
  ) +
  ggtitle("nifH Phylum Relative Abundance by Site") +
```

```
ylab("Relative Abundance (%)")+
scale_x_discrete(labels = c("Conrad", "Corvallis", "Huntley Dryland",
                           "Huntley Irrigated", "Kalispell", "Moccasin",
                           "Sidney Dryland", "Sidney Irrigated", "Richland"))+
theme(axis.text.x = element_text(angle = 90, hjust = 1),
      axis.text.y = element_blank(), panel.background = element_blank())
```



Publish plot to tiff

When published bars lose little white lines

```
## pdf
## 2
```

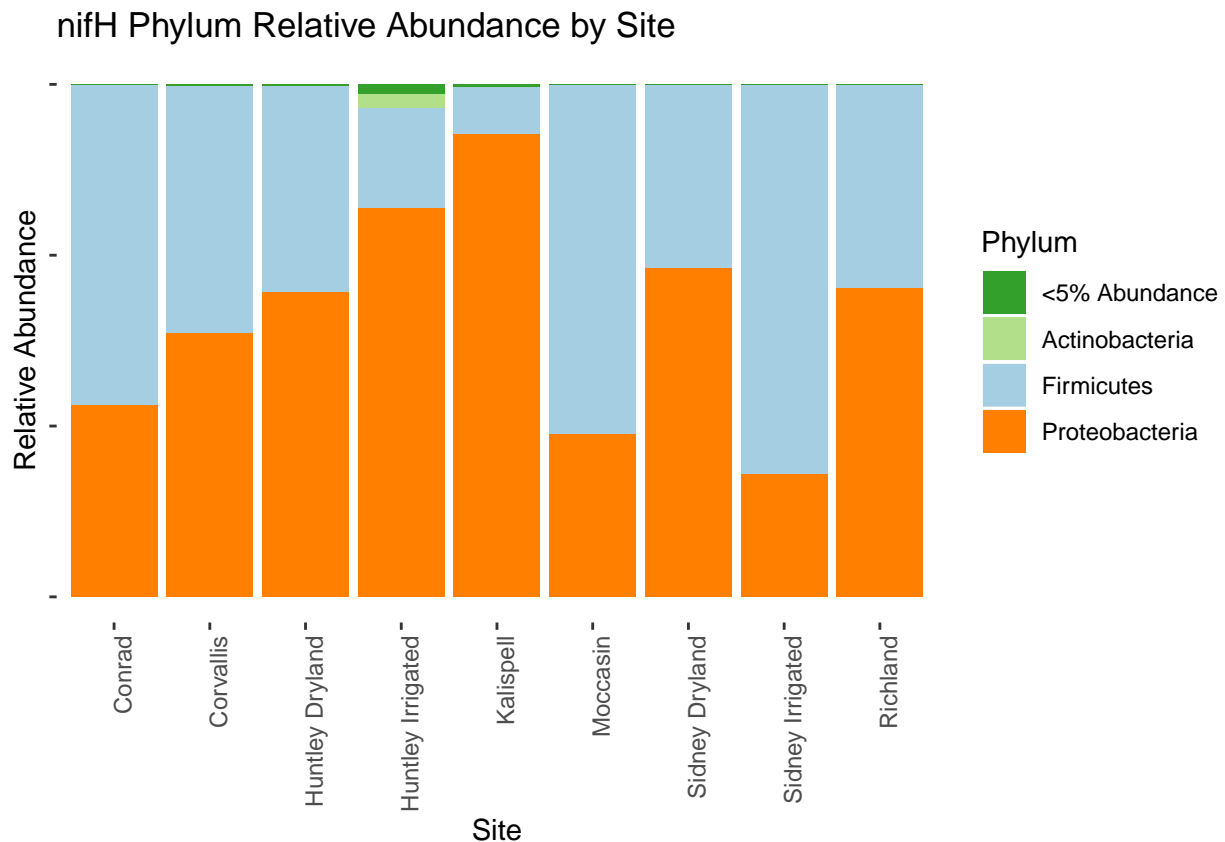
Different Colors

```
ggplot(data = data_nifH_phylum, aes(x = Site, y = Abundance, fill = phylum))+
geom_bar(aes(fill = phylum), stat = "identity", position = "stack", show.legend = TRUE)+
scale_fill_manual(name = "Phylum",
                  values=c( "#33a02c", "#b2df8a", "#a6cee3", "#ff7f00"),
                  breaks=c("p__proteobacteria", "p__firmicutes",
                           "p__actinobacteria", "<5% abund" ),
                  labels = c( "Proteobacteria", "Firmicutes",
                              "Actinobacteria", "<5% Abundance"),
                  guide = guide_legend(reverse = TRUE))
```

```

    )+
ggtitle("nifH Phylum Relative Abundance by Site")+
ylab("Relative Abundance")+
scale_x_discrete(labels = c("Conrad", "Corvallis",
                            "Huntley Dryland", "Huntley Irrigated", "Kalispell",
                            "Moccasin", "Sidney Dryland", "Sidney Irrigated", "Richland"))+
theme(axis.text.x = element_text(angle = 90, hjust = 1),
      axis.text.y = element_blank(), panel.background = element_blank())

```



nifH genus barplot

```

physeq_nifH_ord_1 = transform_sample_counts(physeq_nifH_ord, function(x) x / sum(x) )
physeq_nifH_ord_genus <- tax_glom(physeq_nifH_ord_1, "genus")
data_nifH_genus <- psmelt(physeq_nifH_ord_genus)
data_nifH_genus$genus<-as.character(data_nifH_genus$genus)
data_nifH_genus$genus[data_nifH_genus$Abundance<0.1]<-"<10% abund"
unique(data_nifH_genus$genus)

```

```

## [1] "g__paenibacillus"      "g__rhizobium"          "g__geobacter"
## [4] "g__rhodopseudomonas"  "g__bradyrhizobium"     "g__gluconacetobacter"
## [7] "g__dechloromonas"     "g__agrobacterium"      "g__arthrobacter"
## [10] "<10% abund"

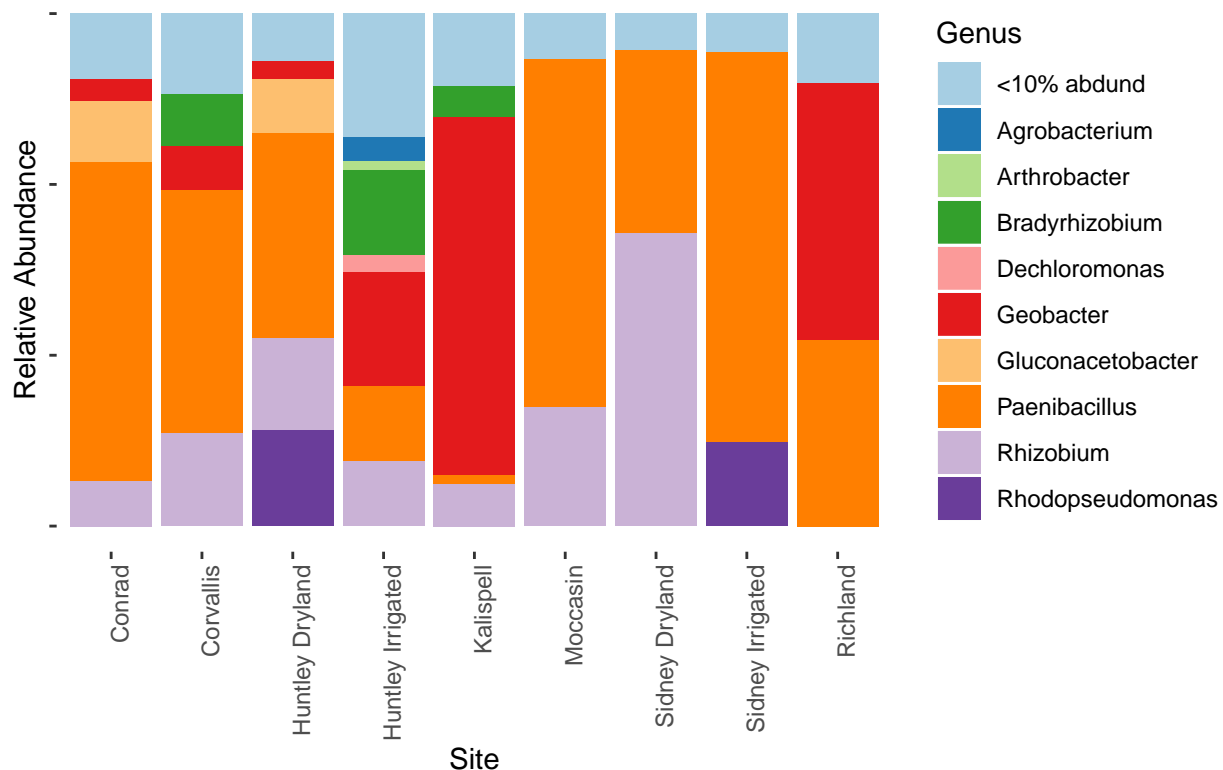
```

```

ggplot(data = data_nifH_genus, aes(x = Site, y = Abundance, fill = genus))+
geom_bar(aes(fill = genus), stat = "identity", position = "stack", show.legend = TRUE)+
scale_fill_manual(name = "Genus",
  values=c('#a6cee3','#1f78b4','#b2df8a','#33a02c',
    '#fb9a99','#e31a1c','#fdbf6f','#ff7f00',
    '#cab2d6','#6a3d9a','#ffff99','#b15928'),
  breaks=c("<10% abund", "g__agrobacterium", "g__arthrobacter",
    "g__bradyrhizobium", "g__dechloromonas", "g__geobacter",
    "g__gluconacetobacter", "g__paenibacillus",
    "g__rhizobium", "g__rhodopseudomonas"),
  labels=c("<10% abund", "Agrobacterium", "Arthrobacter",
    "Bradyrhizobium", "Dechloromonas", "Geobacter",
    "Gluconacetobacter", "Paenibacillus",
    "Rhizobium", "Rhodopseudomonas"),
  guide = guide_legend(reverse = FALSE)
)+
ggtitle("nifH Genus Relative Abundance by Site")+
ylab("Relative Abundance")+
scale_x_discrete(labels = c("Conrad", "Corvallis", "Huntley Dryland",
  "Huntley Irrigated", "Kalispell", "Moccasin",
  "Sidney Dryland", "Sidney Irrigated", "Richland"))+
theme(axis.text.x = element_text(angle = 90, hjust = 1), axis.text.y = element_blank(),
  panel.background = element_blank())

```

nifH Genus Relative Abundance by Site



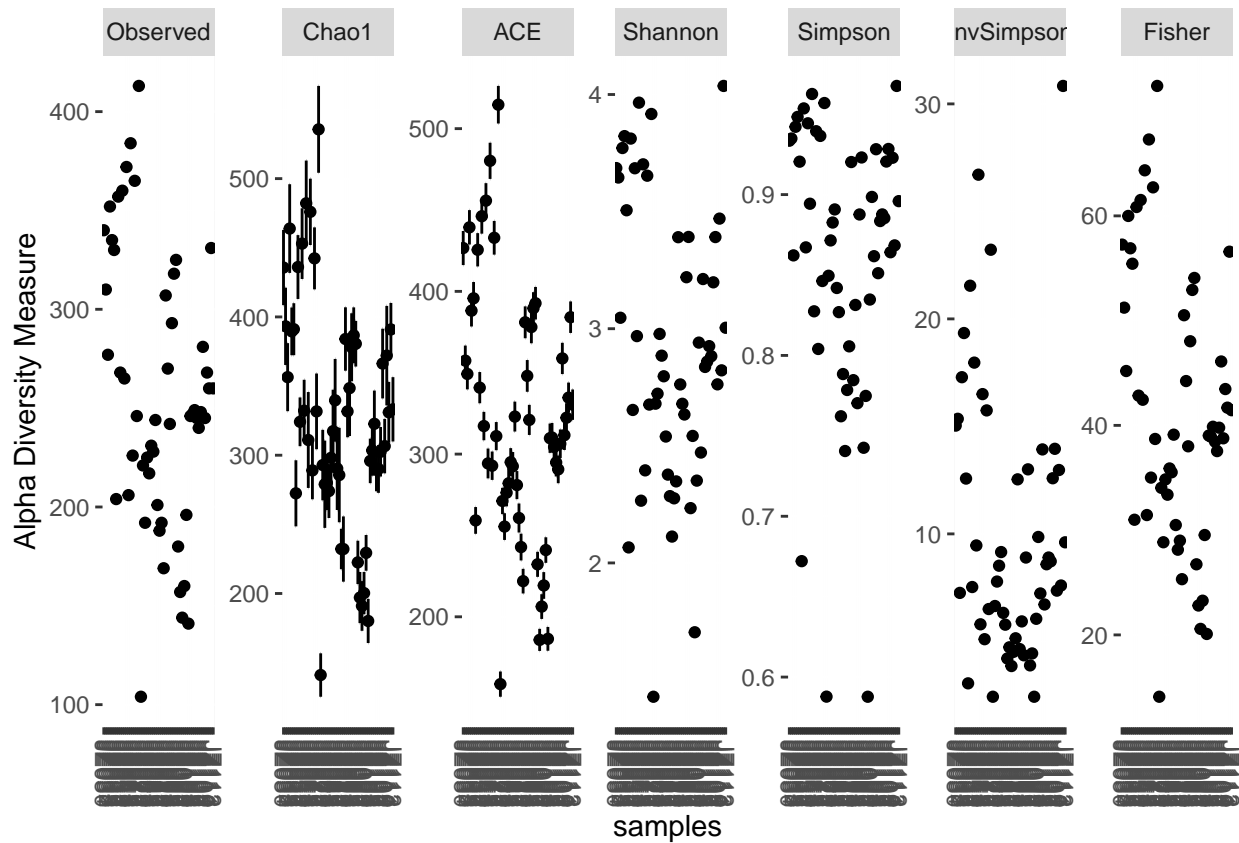
Publish to a tiff image


```
## pdf
## 2
```

Alpha diversity metrics

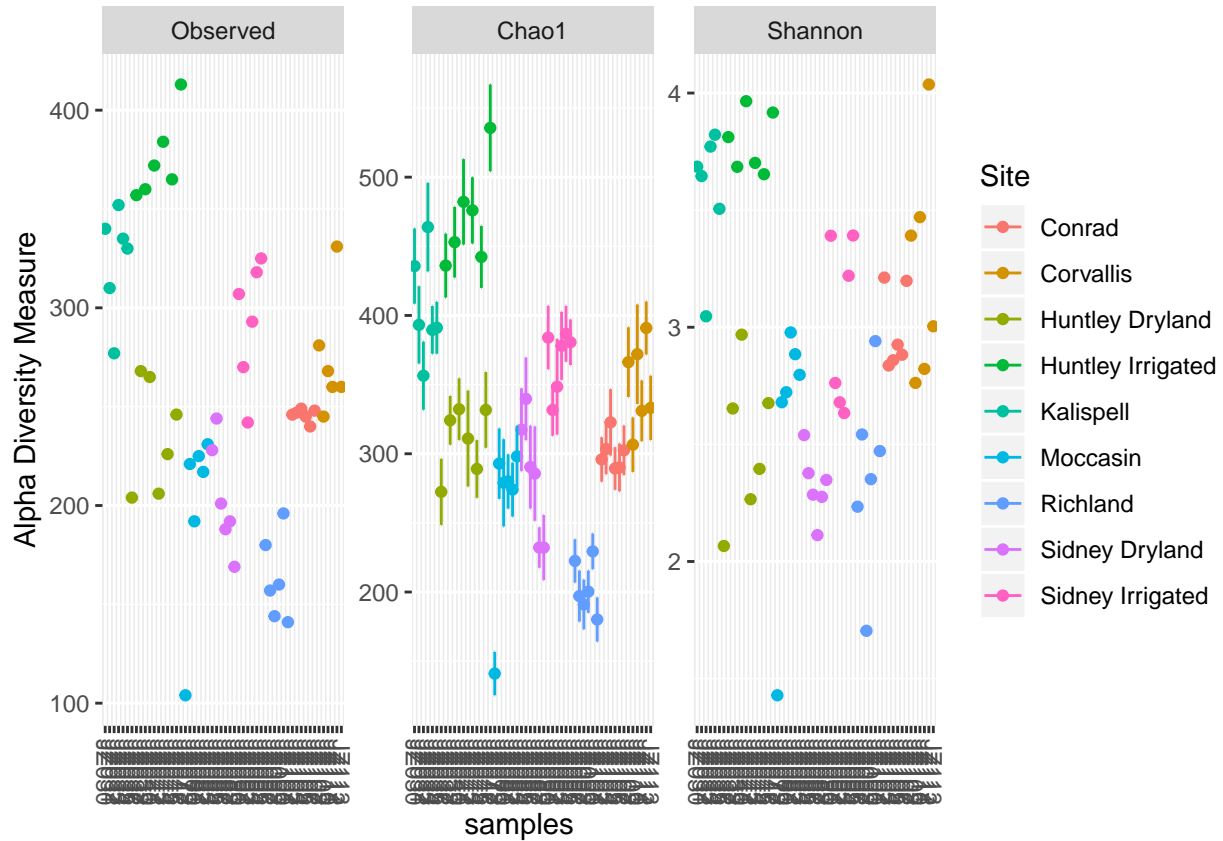
Use phyloseq internal packages to calculate the alpha diversity

```
plot_richness(physeq_nifH_trim)
```



Simplify to just Observed, Chao1, and Shannon

```
plot_richness(physeq_nifH_trim, measures = c("Observed", "Chao1", "Shannon"), color = "Site")
```



```
rich_nifH<-estimate_richness(physeq_nifH_trim, split = TRUE)
```

```
#First merge data sets with meta2
meta2$sample_names<-rownames(meta2)
rich_nifH$sample_names<-rownames(rich_nifH)
meta_nifH<-merge(meta2, rich_nifH, by = "sample_names")
rownames(meta_nifH)<-meta_nifH$sample_names
meta_nifH<-meta_nifH[,-1]
head(meta_nifH)[,49:54]
```

	ACE	se.ACE	Shannon	Simpson	InvSimpson	Fisher
JZ030	426.6245	10.472146	3.685694	0.9334978	15.037099	57.25766
JZ031	357.4261	9.265222	3.645484	0.9348672	15.353248	51.23836
JZ032	349.2533	9.483790	3.046774	0.8621953	7.256646	45.19042
JZ033	439.3598	10.570712	3.771445	0.9421562	17.287932	59.99256
JZ034	388.1394	9.716513	3.822216	0.9482979	19.341566	56.90461
JZ035	395.7359	10.022588	3.505772	0.9204658	12.573215	55.43243

Make boxplots with Observed, Shannon, chao1, and inverse simpson

```
#use ggpubr for plot
nifH_Observ<-ggboxplot(meta_nifH, x = "Site", y = "Observed",
  rug = TRUE,
```

```

fill = "Site", xlab = " ",
palette = farm_col_paired)+
rremove("x.text")

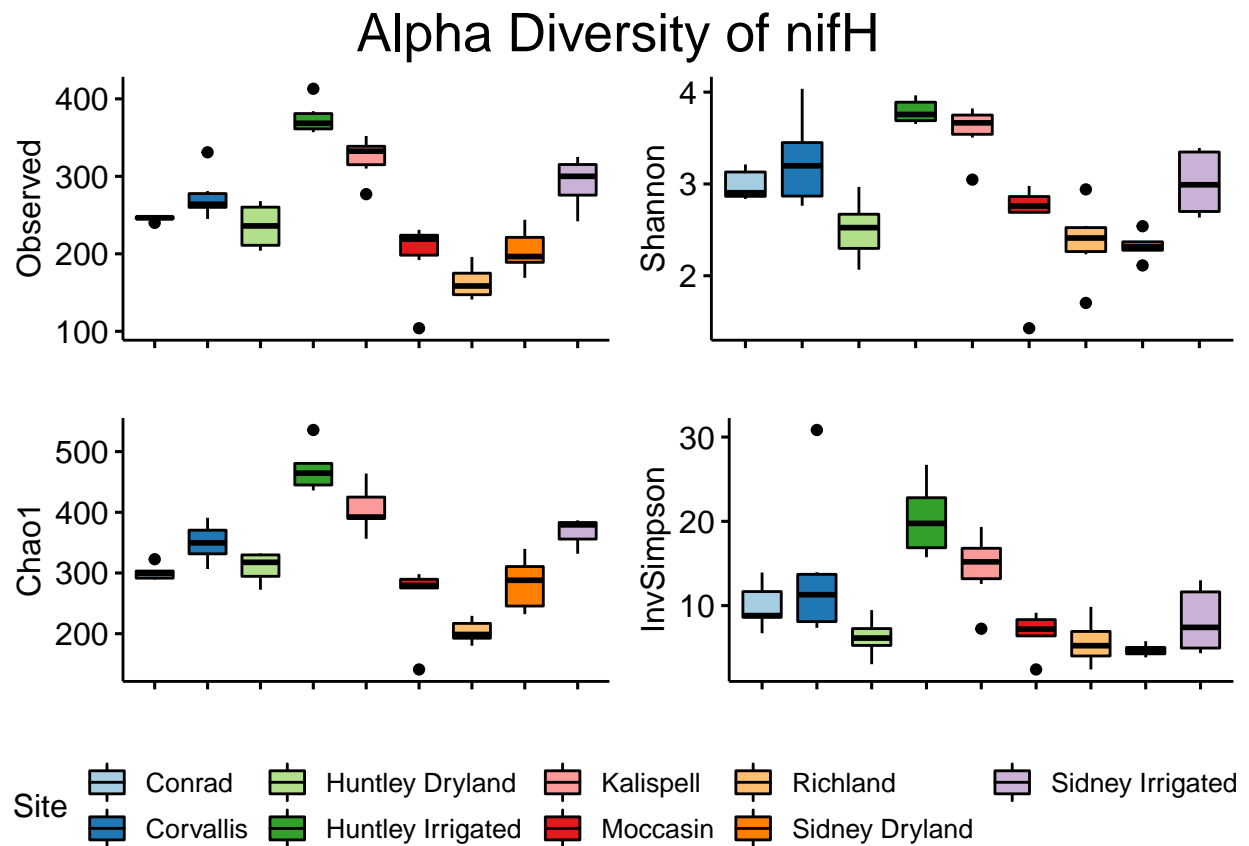
nifH_Shannon<-ggboxplot(meta_nifH, x = "Site", y = "Shannon",
  rug = TRUE,
  fill = "Site", xlab = " ",
  palette = farm_col_paired)+
rremove("x.text")

nifH_Chao<- ggboxplot(meta_nifH, x = "Site", y = "Chao1",
  rug = TRUE,
  fill = "Site", xlab = " ",
  palette = farm_col_paired)+
rremove("x.text")

nifH_InvSim<- ggboxplot(meta_nifH, x = "Site", y = "InvSimpson",
  rug = TRUE,
  fill = "Site", xlab = " ",
  palette = farm_col_paired)+
rremove("x.text")

alpha_nifH_fig<-ggarrange(nifH_Observ, nifH_Shannon, nifH_Chao, nifH_InvSim, ncol = 2, nrow = 2, common
  annotate_figure(alpha_nifH_fig, top = text_grob("Alpha Diversity of nifH ", size = 20))

```



Publish to .tiff

```
## pdf
## 2
```

Observed- total observed OTUs

Chao1- estimate diversity and assumes that the number of observations for a taxa has a Poisson distribution and corrects for variance

Shannon- Number of OTUs (richness) scaled to the evenness

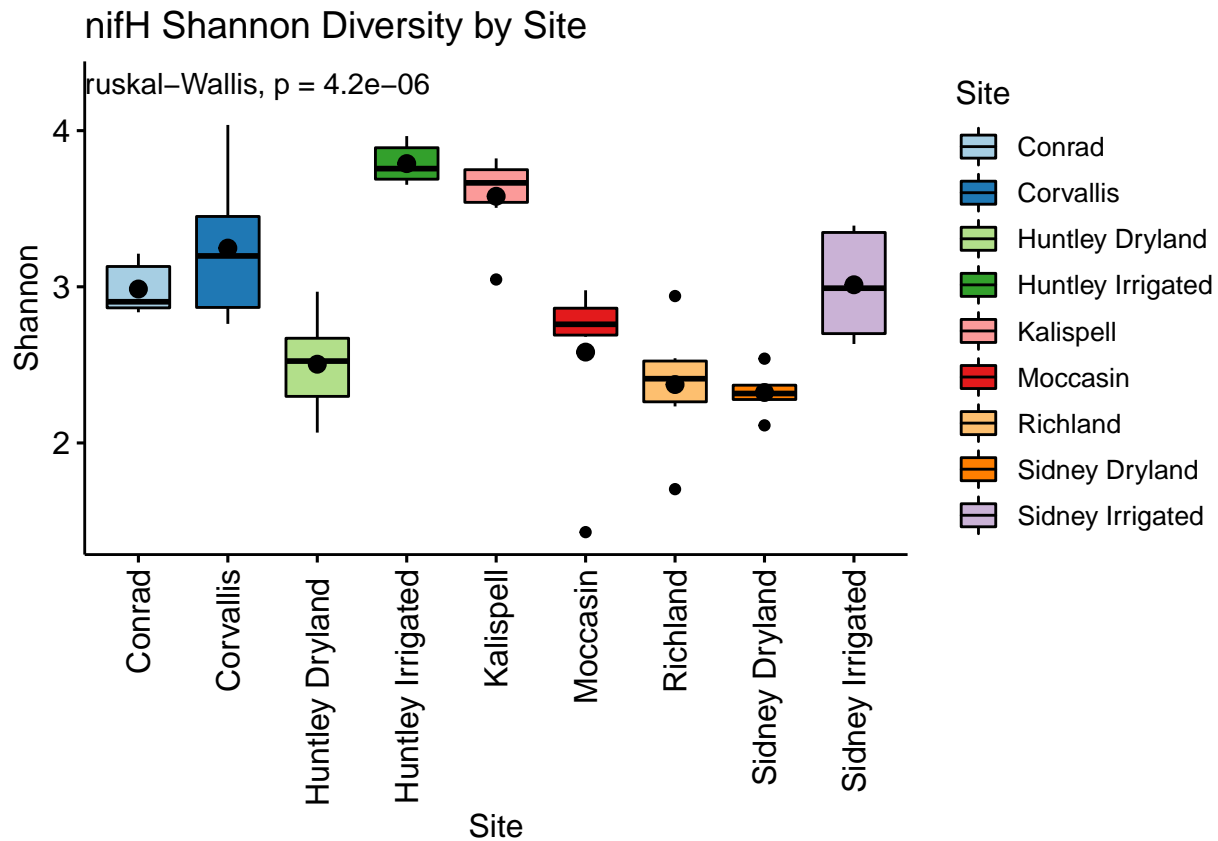
Simpson- scale of dominance probability of any two individuals drawn at random belonging to the same species

Source: <http://biology.kenyon.edu/courses/biol229/diversity.pdf>

Just shannon plot by site

```
#use ggpubr for plot
ggboxplot(meta_nifH, x = "Site", y = "Shannon",
  add = "mean", rug = TRUE,
  fill = "Site",
  title = "nifH Shannon Diversity by Site", palette = farm_col_paired, legend = "right")+
  stat_compare_means(label.y = 4.3, p.adjust.method = "bonferroni")+
  rotate_x_text()
```

```
## Warning: Ignoring unknown parameters: p.adjust.method
```



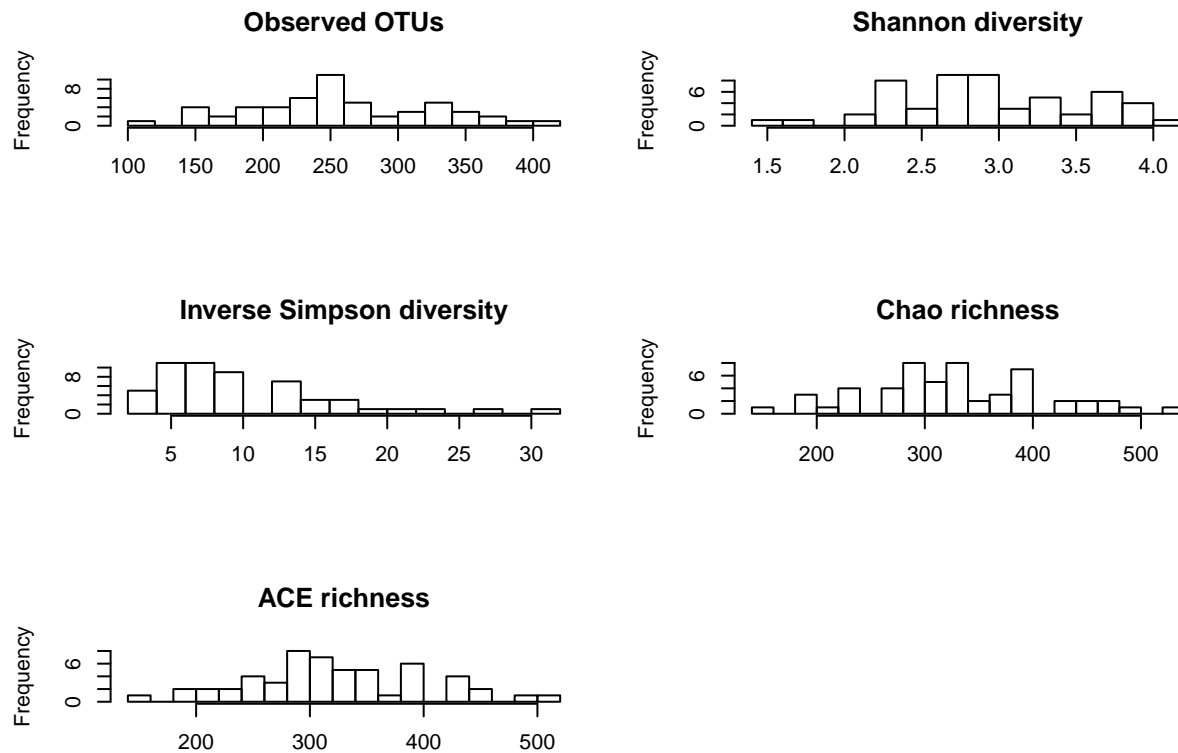
Determine if distribution is normal for each diversity Metric

Used the following protocol :

https://rpubs.com/dillmcfarlan/R_microbiotaSOP

```
#Create 2x2 plot environment so that we can see all 4 metrics at once.
par(mfrow = c(3, 2))

#Then plot each metric.
hist(rich_nifH$Observed, main="Observed OTUs", xlab="", breaks=15)
hist(rich_nifH$Shannon, main="Shannon diversity", xlab="", breaks=10)
hist(rich_nifH$InvSimpson, main="Inverse Simpson diversity", xlab="", breaks=15)
hist(rich_nifH$Chao1, main="Chao richness", xlab="", breaks=15)
hist(rich_nifH$ACE, main="ACE richness", xlab="", breaks=15)
```



Test for normalcy using the shapiro test. The null hypothesis for this test is that the data are normally distributed, if the p-value is greater than 0.05, then the null hypothesis is not rejected.

```
shapiro.test(rich_nifH$Observed)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  rich_nifH$Observed
## W = 0.98368, p-value = 0.6687
```

```
shapiro.test(rich_nifH$Shannon)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  rich_nifH$Shannon
## W = 0.97519, p-value = 0.3226
```

```
shapiro.test(rich_nifH$InvSimpson)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  rich_nifH$InvSimpson
## W = 0.87892, p-value = 5.746e-05
```

```
shapiro.test(rich_nifH$Chao1)
```

```
##  
## Shapiro-Wilk normality test  
##  
## data: rich_nifH$Chao1  
## W = 0.98578, p-value = 0.768
```

```
shapiro.test(rich_nifH$ACE)
```

```
##  
## Shapiro-Wilk normality test  
##  
## data: rich_nifH$ACE  
## W = 0.98743, p-value = 0.8402
```

Use ANOVA on alpha diversity metrics for Shannon variables because the other are not normal

Shannon anova

```
aov_shannon_site_nifH <- aov(Shannon ~ Site, meta_nifH)  
summary(aov_shannon_site_nifH)
```

```
##           Df Sum Sq Mean Sq F value    Pr(>F)  
## Site           8 13.506   1.6882    13.67 8.29e-10 ***  
## Residuals     45  5.558   0.1235  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Site location has a significant effect on Shannon diversity

Correct for multiple comparisons

```
shannon_nifH_site <- TukeyHSD(aov_shannon_site_nifH, "Site", ordered = TRUE)  
shannon_nifH_site
```

```
## Tukey multiple comparisons of means  
## 95% family-wise confidence level  
## factor levels have been ordered  
##  
## Fit: aov(formula = Shannon ~ Site, data = meta_nifH)  
##  
## $Site  
##           diff           lwr           upr  
## Richland-Sidney Dryland    0.05089297 -0.609986678 0.7117726  
## Huntley Dryland -Sidney Dryland 0.18109141 -0.479788234 0.8419711  
## Moccasin-Sidney Dryland    0.25873022 -0.402149425 0.9196099
```

## Conrad-Sidney Dryland	0.66272066	0.001841014	1.3236003
## Sidney Irrigated-Sidney Dryland	0.68988515	0.029005499	1.3507648
## Corvallis-Sidney Dryland	0.92476069	0.263881042	1.5856403
## Kalispell-Sidney Dryland	1.25652759	0.595647944	1.9174072
## Huntley Irrigated-Sidney Dryland	1.46594982	0.805070172	2.1268295
## Huntley Dryland -Richland	0.13019844	-0.530681204	0.7910781
## Moccasin-Richland	0.20783725	-0.453042395	0.8687169
## Conrad-Richland	0.61182769	-0.049051955	1.2727073
## Sidney Irrigated-Richland	0.63899218	-0.021887471	1.2998718
## Corvallis-Richland	0.87386772	0.212988073	1.5347474
## Kalispell-Richland	1.20563462	0.544754974	1.8665143
## Huntley Irrigated-Richland	1.41505685	0.754177202	2.0759365
## Moccasin-Huntley Dryland	0.07763881	-0.583240839	0.7385185
## Conrad-Huntley Dryland	0.48162925	-0.179250399	1.1425089
## Sidney Irrigated-Huntley Dryland	0.50879373	-0.152085915	1.1696734
## Corvallis-Huntley Dryland	0.74366928	0.082789629	1.4045489
## Kalispell-Huntley Dryland	1.07543618	0.414556530	1.7363158
## Huntley Irrigated-Huntley Dryland	1.28485841	0.623978758	1.9457381
## Conrad-Moccasin	0.40399044	-0.256889209	1.0648701
## Sidney Irrigated-Moccasin	0.43115492	-0.229724724	1.0920346
## Corvallis-Moccasin	0.66603047	0.005150819	1.3269101
## Kalispell-Moccasin	0.99779737	0.336917720	1.6586770
## Huntley Irrigated-Moccasin	1.20721960	0.546339948	1.8680992
## Sidney Irrigated-Conrad	0.02716448	-0.633715164	0.6880441
## Corvallis-Conrad	0.26204003	-0.398839620	0.9229197
## Kalispell-Conrad	0.59380693	-0.067072719	1.2546866
## Huntley Irrigated-Conrad	0.80322916	0.142349509	1.4641088
## Corvallis-Sidney Irrigated	0.23487554	-0.426004105	0.8957552
## Kalispell-Sidney Irrigated	0.56664244	-0.094237204	1.2275221
## Huntley Irrigated-Sidney Irrigated	0.77606467	0.115185024	1.4369443
## Kalispell-Corvallis	0.33176690	-0.329112747	0.9926465
## Huntley Irrigated-Corvallis	0.54118913	-0.119690519	1.2020688
## Huntley Irrigated-Kalispell	0.20942223	-0.451457420	0.8703019
##	p adj		
## Richland-Sidney Dryland	0.9999994		
## Huntley Dryland -Sidney Dryland	0.9923049		
## Moccasin-Sidney Dryland	0.9333949		
## Conrad-Sidney Dryland	0.0488792		
## Sidney Irrigated-Sidney Dryland	0.0347461		
## Corvallis-Sidney Dryland	0.0012087		
## Kalispell-Sidney Dryland	0.0000055		
## Huntley Irrigated-Sidney Dryland	0.0000002		
## Huntley Dryland -Richland	0.9992243		
## Moccasin-Richland	0.9813852		
## Conrad-Richland	0.0893804		
## Sidney Irrigated-Richland	0.0651612		
## Corvallis-Richland	0.0026283		
## Kalispell-Richland	0.0000128		
## Huntley Irrigated-Richland	0.0000004		
## Moccasin-Huntley Dryland	0.9999841		
## Conrad-Huntley Dryland	0.3228127		
## Sidney Irrigated-Huntley Dryland	0.2557929		
## Corvallis-Huntley Dryland	0.0170802		
## Kalispell-Huntley Dryland	0.0001098		


```
## Huntley Irrigated-Huntley Dryland 0.0000034
## Conrad-Moccasin 0.5572599
## Sidney Irrigated-Moccasin 0.4701256
## Corvallis-Moccasin 0.0469200
## Kalispell-Moccasin 0.0003838
## Huntley Irrigated-Moccasin 0.0000125
## Sidney Irrigated-Conrad 1.0000000
## Corvallis-Conrad 0.9286970
## Kalispell-Conrad 0.1093235
## Huntley Irrigated-Conrad 0.0074286
## Corvallis-Sidney Irrigated 0.9612154
## Kalispell-Sidney Irrigated 0.1461472
## Huntley Irrigated-Sidney Irrigated 0.0109195
## Kalispell-Corvallis 0.7805875
## Huntley Irrigated-Corvallis 0.1888886
## Huntley Irrigated-Kalispell 0.9804879
```

Write to table

Plot Irrigation managment against diffrent diveristies

```
nifH_irr_Observ<-ggboxplot(meta_nifH, x = "Plot", y = "Observed",
  rug = TRUE,
  fill = "Plot", xlab = " ",
  palette = farm_col_paired)+
  rremove("x.text")

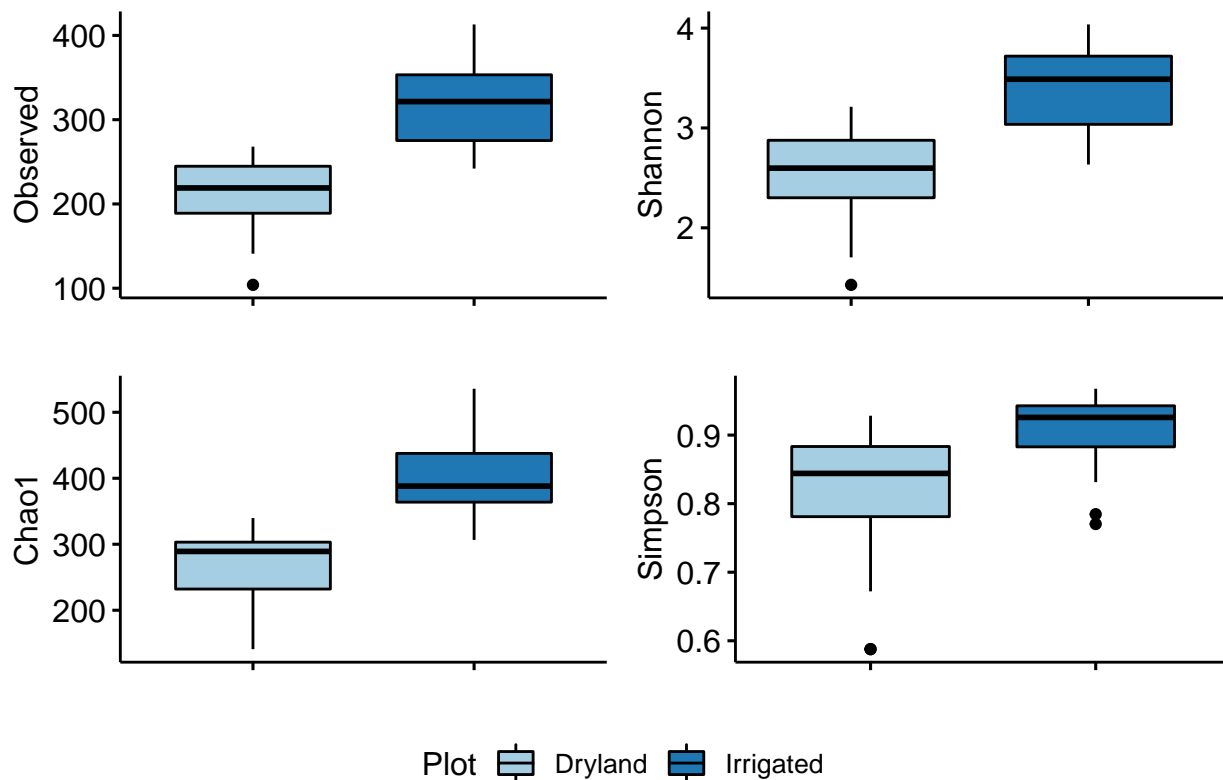
nifH_irr_Shannon<-ggboxplot(meta_nifH, x = "Plot", y = "Shannon",
  rug = TRUE,
  fill = "Plot", xlab = " ",
  palette = farm_col_paired)+
  rremove("x.text")

nifH_irr_Chao<- ggboxplot(meta_nifH, x = "Plot", y = "Chao1",
  rug = TRUE,
  fill = "Plot", xlab = " ",
  palette = farm_col_paired)+
  rremove("x.text")

nifH_irr_InvSim<- ggboxplot(meta_nifH, x = "Plot", y = "Simpson",
  rug = TRUE,
  fill = "Plot", xlab = " ",
  palette = farm_col_paired)+
  rremove("x.text")

alpha_nifH_irr_fig<-ggarrange(nifH_irr_Observ, nifH_irr_Shannon, nifH_irr_Chao, nifH_irr_InvSim, ncol =
  annotate_figure(alpha_nifH_irr_fig, top = text_grob("Alpha Diversity of nifH by Irrigation", size = 20))
```

Alpha Diversity of nifH by Irrigation



Check for significance in the other metrics to for irrigation (Dryland vs irrigated)

Since Shannon was the only normally distributed data set I will use Kruskal Wallis test for others

```
KW_observed_irr_nifH <- kruskal.test(Observed ~ Plot, meta_nifH)
KW_observed_irr_nifH
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  Observed by Plot
## Kruskal-Wallis chi-squared = 34.831, df = 1, p-value = 3.596e-09
```

```
aov_observed_irr_nifH <- aov(Observed ~ Plot, meta_nifH)
summary(aov_observed_irr_nifH)
```

```
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Plot         1 153010   153010      81 3.47e-12 ***
## Residuals    52   98225     1889
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Significance in observed

```
aov_observed_irr_nifH <- aov(Observed ~ Plot/Site, meta_nifH)
summary(aov_observed_irr_nifH)
```

```
##              Df Sum Sq Mean Sq F value    Pr(>F)
## Plot           1 153010   153010   185.99 < 2e-16 ***
## Plot:Site       7  61203     8743    10.63 8.31e-08 ***
## Residuals      45  37021      823
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Plot is significant but the interaction of Plot:Site is also significant so we cannot say there is actual correlation

```
aov_Chao1_irr_nifH <- aov(Chao1 ~ Plot, meta_nifH)
summary(aov_Chao1_irr_nifH)
```

```
##              Df Sum Sq Mean Sq F value    Pr(>F)
## Plot           1 214872   214872     75 1.17e-11 ***
## Residuals      52 148968     2865
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Significant in chao1

```
aov_Chao1_irr_nifH <- aov(Chao1 ~ Plot/Site, meta_nifH)
summary(aov_Chao1_irr_nifH)
```

```
##              Df Sum Sq Mean Sq F value    Pr(>F)
## Plot           1 214872   214872   177.83 < 2e-16 ***
## Plot:Site       7  94595   13514    11.18 4.22e-08 ***
## Residuals      45  54373    1208
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
aov_Simpson_irr_nifH <- aov(Simpson ~ Plot, meta_nifH)
summary(aov_Simpson_irr_nifH)
```

```
##              Df Sum Sq Mean Sq F value    Pr(>F)
## Plot           1  0.1044  0.10442    19.29 5.55e-05 ***
## Residuals      52  0.2815  0.00541
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
aov_Simpson_irr_nifH <- aov(Simpson ~ Plot/Site, meta_nifH)
summary(aov_Simpson_irr_nifH)
```

```
##              Df Sum Sq Mean Sq F value    Pr(>F)
## Plot          1 0.10442 0.10442   23.02 1.79e-05 ***
## Plot:Site      7 0.07744 0.01106    2.44  0.033  *
## Residuals     45 0.20407 0.00453
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

All alpha diversities are significant with irrigation method but they are nested in Site so we must look within site (Huntley and Sidney) later. It is hard to prove anything in the statewide survey with univariate ANOVAs

```
aov_shannon_pea_nifH <- aov(Shannon ~ Pea_variety + Site, meta_nifH)
summary(aov_shannon_pea_nifH)
```

```
##              Df Sum Sq Mean Sq F value    Pr(>F)
## Pea_variety   6  0.898  0.1497   1.178   0.338
## Site          8 13.209  1.6511  12.991 6.73e-09 ***
## Residuals     39  4.957  0.1271
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
aov_shannon_prevcrop_nifh<-aov(Shannon ~ prev_crop, meta_nifH)
summary(aov_shannon_prevcrop_nifh)
```

```
##              Df Sum Sq Mean Sq F value    Pr(>F)
## prev_crop      3 10.965   3.655  22.57 2.23e-09 ***
## Residuals     50  8.098   0.162
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
aov_shannon_prevcrop_site_nifh<-aov(Shannon ~ prev_crop/Site, meta_nifH)
summary(aov_shannon_prevcrop_site_nifh)
```

```
##              Df Sum Sq Mean Sq F value    Pr(>F)
## prev_crop      3 10.965   3.655  29.594 1.01e-10 ***
## prev_crop:Site  5  2.540   0.508   4.114 0.00367 **
## Residuals     45  5.558   0.124
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
aov_shannon_till_nifh<-aov(Shannon ~ Tillage, meta_nifH)
summary(aov_shannon_till_nifh)
```

```
##              Df Sum Sq Mean Sq F value    Pr(>F)
## Tillage        2  0.844  0.4220   1.181  0.315
## Residuals     51 18.219  0.3572
```

```
aov_shannon_till_site_nifh<-aov(Shannon ~ Tillage/Site, meta_nifH)
summary(aov_shannon_till_site_nifh)
```

```
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Tillage      2  0.844   0.4220    3.417   0.0415 *
## Tillage:Site  6 12.662   2.1103   17.086 3.61e-10 ***
## Residuals    45  5.558   0.1235
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

nifH_irr_Observ_scatter<-ggscatter(meta_nifH, x = "total_precip_irr", y = "Observed",
  rug = TRUE,
  add = "reg.line",
  color = "Plot", xlab = " ",
  palette = farm_col_paired)+
  rremove("x.text")

nifH_irr_Shannon_scatter<-ggscatter(meta_nifH, x = "total_precip_irr", y = "Shannon",
  rug = TRUE,
  add = "reg.line",
  color = "Plot", xlab = " ",
  palette = farm_col_paired)+
  rremove("x.text")

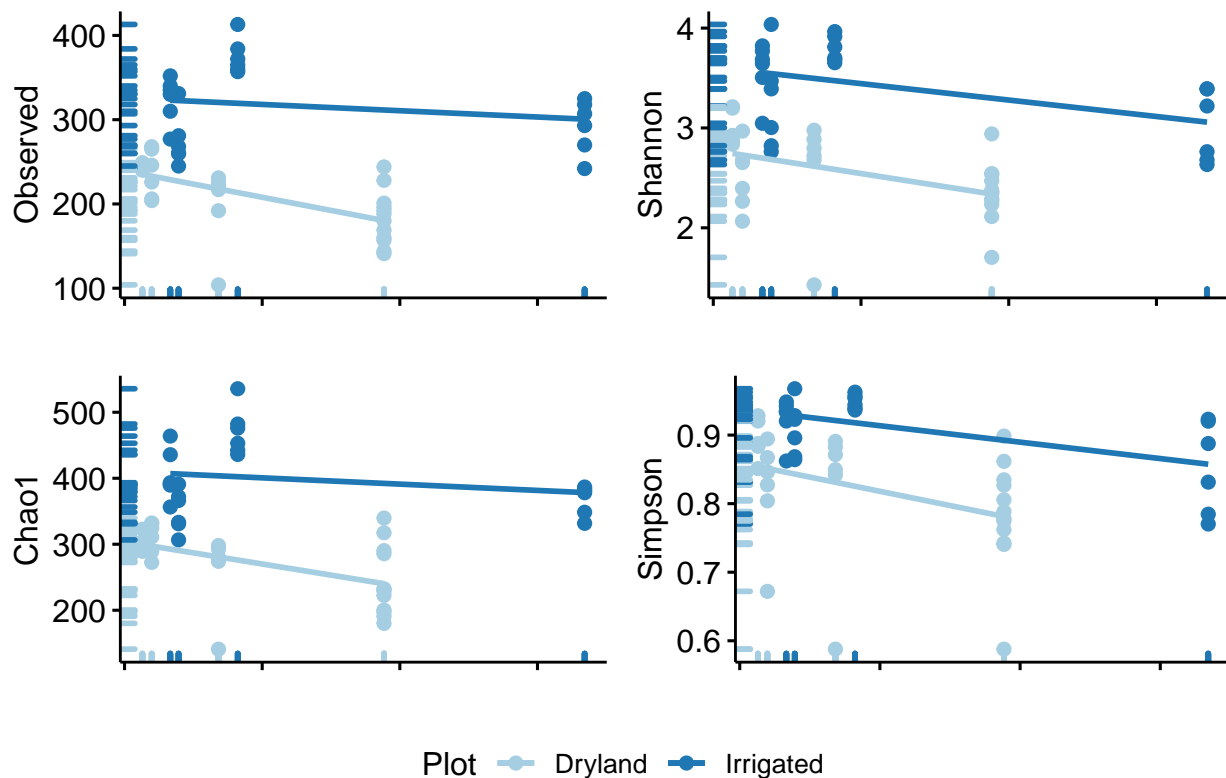
nifH_irr_Chao_scatter<- ggscatter(meta_nifH, x = "total_precip_irr", y = "Chao1",
  rug = TRUE,
  add = "reg.line",
  color = "Plot", xlab = " ",
  palette = farm_col_paired)+
  rremove("x.text")

nifH_irr_InvSim_scatter<- ggscatter(meta_nifH, x = "total_precip_irr", y = "Simpson",
  rug = TRUE,
  add = "reg.line",
  color = "Plot", xlab = " ",
  palette = farm_col_paired)+
  rremove("x.text")

alpha_nifH_irr_fig<-ggarrange(nifH_irr_Observ_scatter, nifH_irr_Shannon_scatter,
  nifH_irr_Chao_scatter, nifH_irr_InvSim_scatter,
  ncol = 2, nrow = 2, common.legend = TRUE, legend = "bottom")

annotate_figure(alpha_nifH_irr_fig,
  top = text_grob("Alpha Diversity of nifH by Irrigation", size = 20))
```

Alpha Diversity of nifH by Irrigation



Plot ordination

Used the following protocol from the phyloseq tutorial:

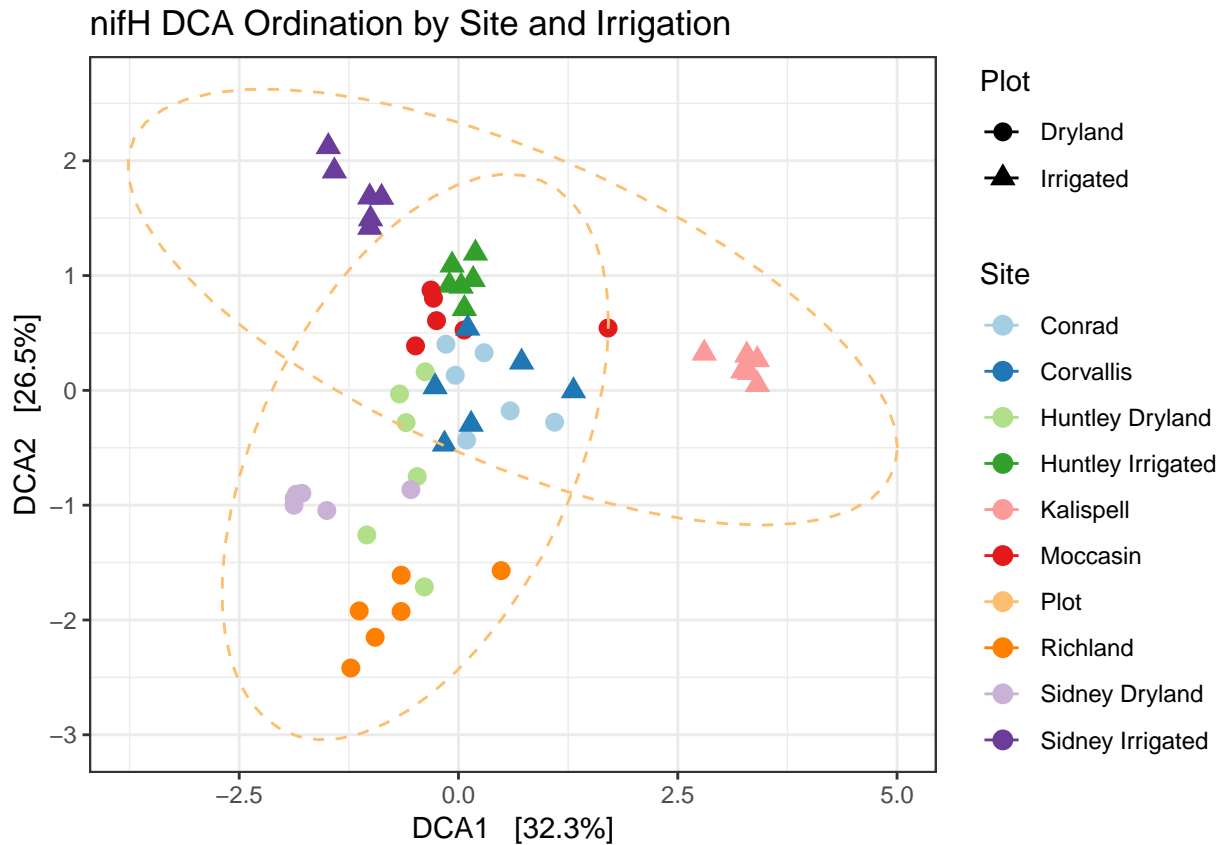
https://joey711.github.io/phyloseq/plot_ordination-examples.html

Resources for ordination statistics

http://ordination.okstate.edu/overview.htm#Principal_Components_Analysis Application of multivariate statistical techniques in microbial ecology: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4769650/pdf/nihms752912.pdf> Linking multidimensional functional diversity to quantitative methods: a graphical hypothesis-evaluation framework <https://esajournals.onlinelibrary.wiley.com/doi/full/10.1890/15-0688>

DCA Ordination

```
phynifH_ord_DCA <- ordinate(physeq_nifH_ord, "DCA", "bray")
plot_ordination(physeq_nifH_ord, phynifH_ord_DCA, color = "Site", shape = "Plot")+
  geom_point(size = 3)+
  scale_color_manual(values = farm_col_paired)+
  stat_ellipse(type = "norm", linetype = 2, aes(color = "Plot"), show.legend = TRUE) +
  ggtitle("nifH DCA Ordination by Site and Irrigation")+
  theme_bw()
```



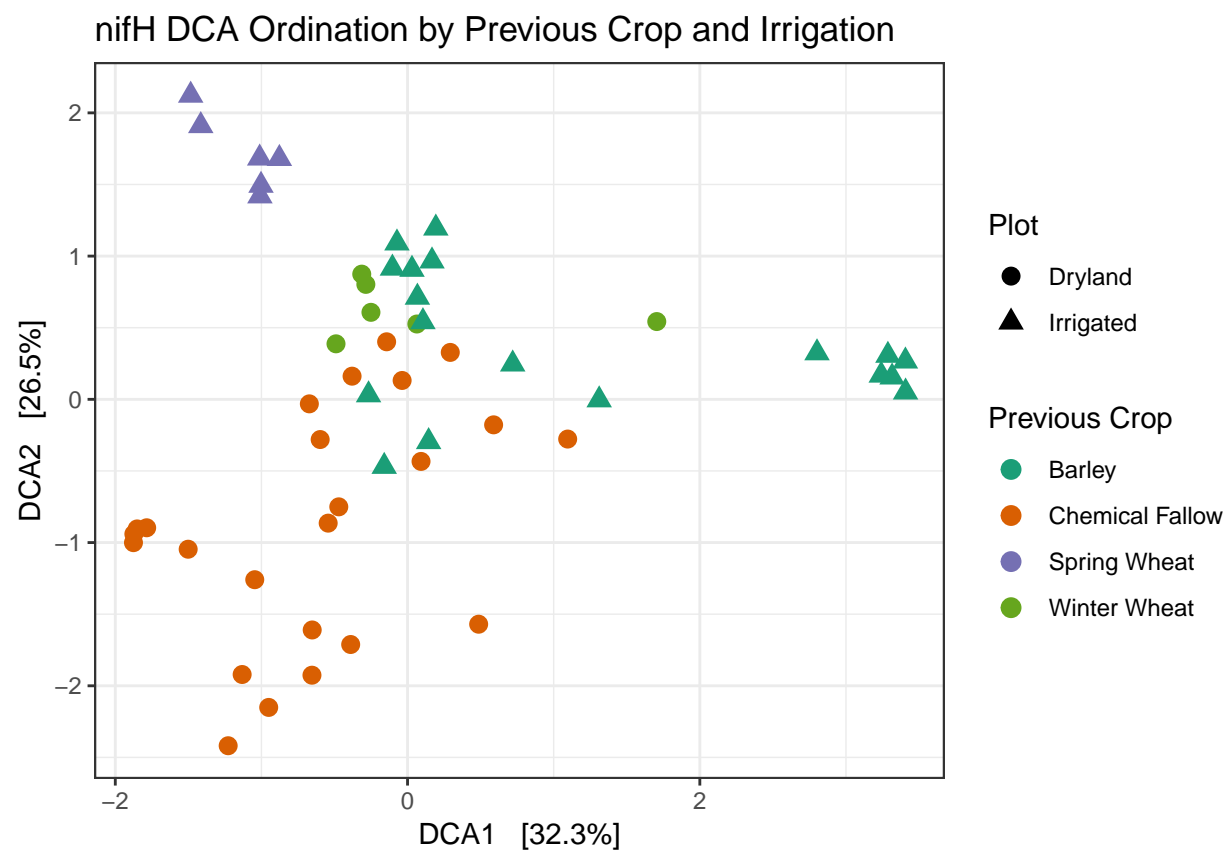
Irrigation is influencing the ordination of the principle components (DCA2 is most likely comprised of irrigation / other farm Management)

Publish to the tiff

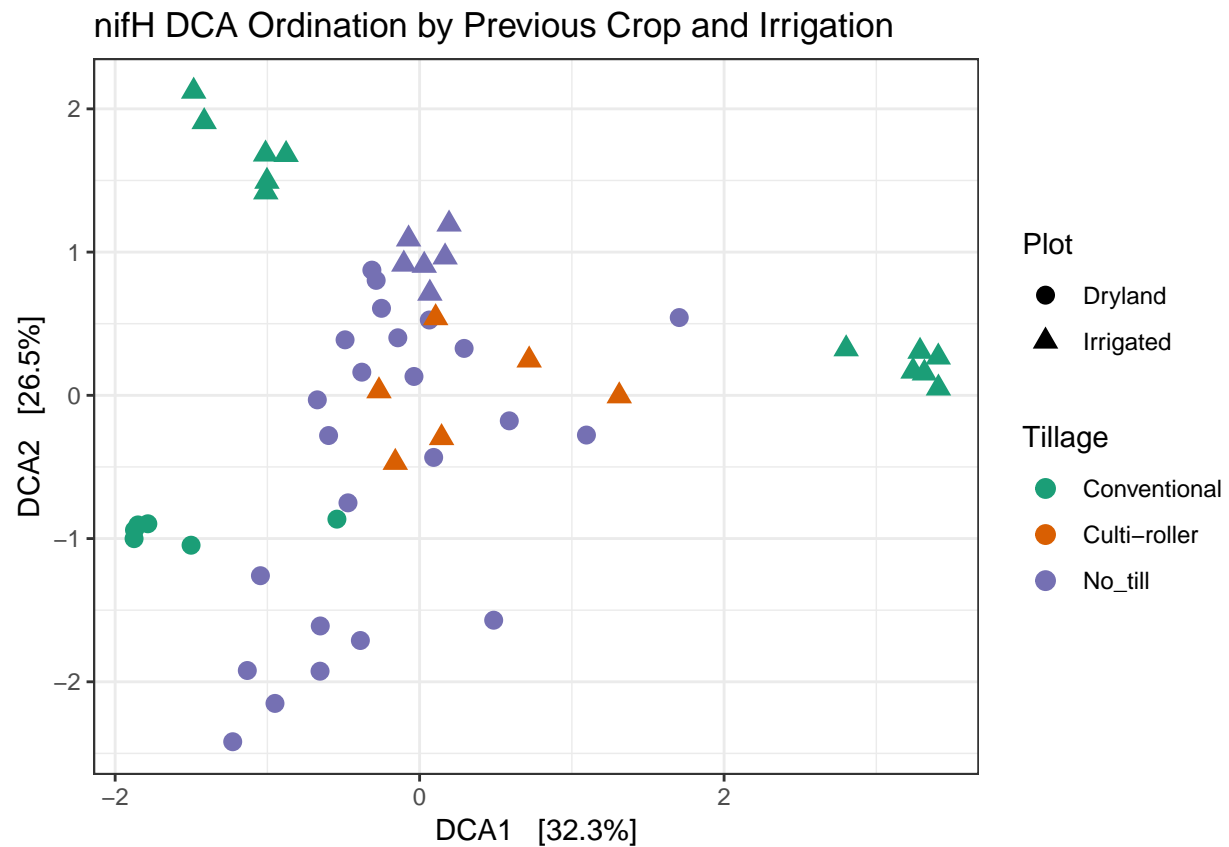
```
## pdf
## 2
```

Will color with other farm management to see if anything is interesting.

```
plot_ordination(physeq_nifH_ord, phynifH_ord_DCA, color = "prev_crop", shape = "Plot")+
  geom_point(size = 3)+
  scale_color_manual(values = c("#1B9E77", "#D95F02", "#7570B3", "#66a61E"),
    name = "Previous Crop",
    breaks=c("barley", "Chem_fallow", "Spring_wheat", "winter_wheat"),
    labels=c("Barley", "Chemical Fallow", "Spring Wheat", "Winter Wheat"))+
  ggtitle("nifH DCA Ordination by Previous Crop and Irrigation")+
  theme_bw()
```



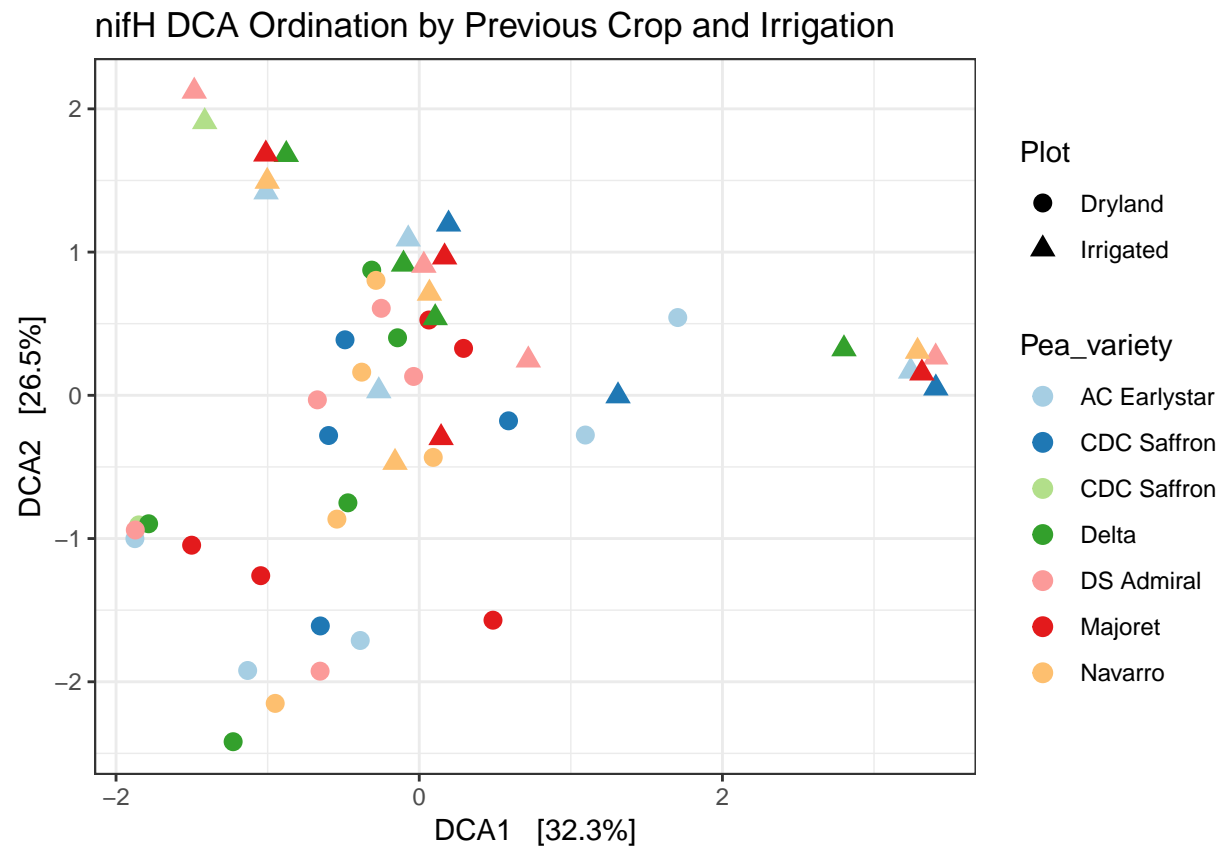
```
plot_ordination(physeq_nifH_ord, phynifH_ord_DCA, color = "Tillage", shape = "Plot")+
  geom_point(size = 3)+
  scale_color_manual(values = farm_col_dark)+
  ggtitle("nifH DCA Ordination by Previous Crop and Irrigation")+
  theme_bw()
```

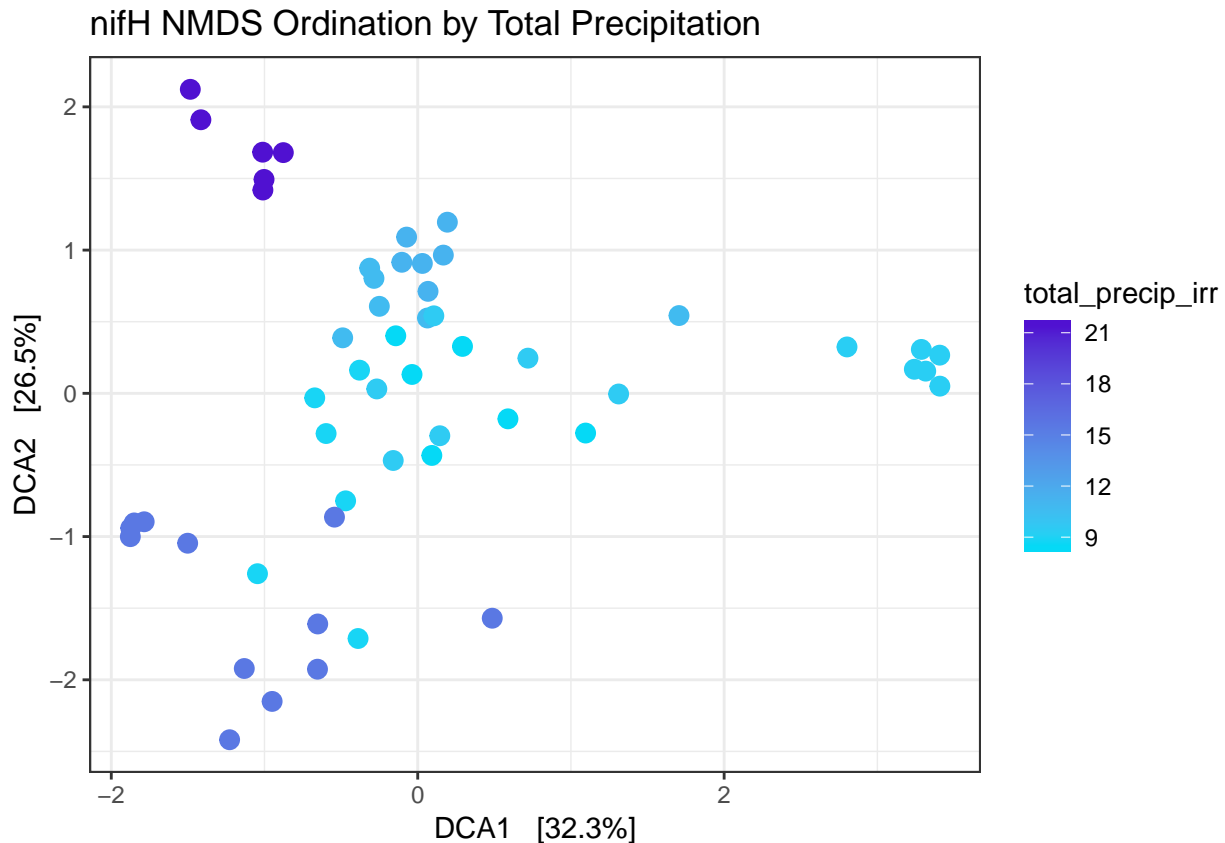
These DCA plot show a distinct separation for Tillage but not as much for previous crop

Pea variety has no correlation or ordnation to bacterial community bray-curtis distance

```
plot_ordination(physeq_nifH_ord, phynifH_ord_DCA, color = "Pea_variety", shape = "Plot")+
  geom_point(size = 3)+
  scale_color_manual(values = farm_col_paired)+
  ggtitle("nifH DCA Ordination by Previous Crop and Irrigation")+
  theme_bw()
```



```
plot_ordination(physeq_nifH_ord, phynifH_ord_DCA, color = "total_precip_irr")+
  geom_point(size = 3)+
  scale_color_gradient(low="#05D9F6", high="#5011D1")+
  ggtitle("nifH NMDS Ordination by Total Precipitation")+
  theme_bw()
```



There might be some Total water gradient effect in the DCA plot

CDA is good but we can also look in nonmetric multidimensional scaling

Contrast between DCA and NMDS

DCA and NMDS are the two most popular methods for indirect gradient analysis. The reason they have remained side-by-side for so long is because, in part, they have different strengths and weaknesses... Some of the issues are relatively minor: for example, computation time is rarely an important consideration, except for the hugest data sets. Some issues are not entirely resolved: the degree to which noise affects NMDS, and the degree to which NMDS finds local rather than global options still need to be determined ... Since NMDS is a distance-based method, all information about species identities is hidden once the distance matrix is created. For many, this is the biggest disadvantage of NMDS... perhaps the biggest difference between the two methods: DCA is based on an underlying model of species distributions, the unimodal model, while NMDS is not. Thus, DCA is closer to a theory of community ecology. However, NMDS may be a method of choice if species composition is determined by factors other than position along a gradient: For example, the species present on islands may have more to do with vicariance biogeography and chance extinction events than with environmental preferences – and for such a system, NMDS would be a better a priori choice. As De'ath (1999) points out, there are two classes of ordination methods - 'species composition restoration' (e.g. NMDS) and 'gradient analysis' (e.g. DCA). The choice between the methods should ultimately be governed by this philosophical distinction. - http://ordination.okstate.edu/overview.htm#Principal_Components_Analysis

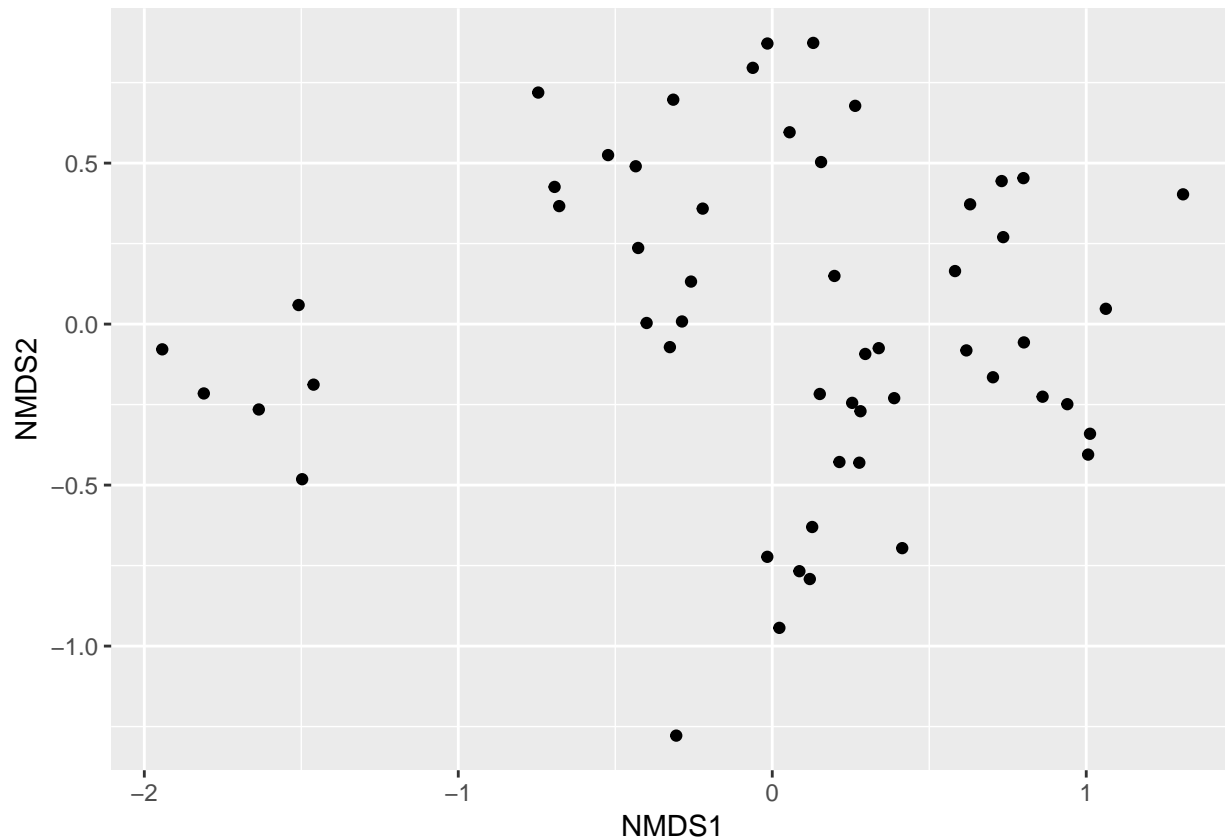
NMDS might be a better choice since we have non gradient determining facotrs site and farm managment effecting the bacteria community

NMDS ordination

```
phynifH_ord_NMDS <- ordinate(physeq_nifH_ord, "NMDS", "bray")
```

```
## Square root transformation
## Wisconsin double standardization
## Run 0 stress 0.1929531
## Run 1 stress 0.1866013
## ... New best solution
## ... Procrustes: rmse 0.0548474  max resid 0.1690312
## Run 2 stress 0.192952
## Run 3 stress 0.201325
## Run 4 stress 0.2186671
## Run 5 stress 0.1929529
## Run 6 stress 0.192428
## Run 7 stress 0.1866138
## ... Procrustes: rmse 0.002278817  max resid 0.009262113
## ... Similar to previous best
## Run 8 stress 0.1924281
## Run 9 stress 0.2013283
## Run 10 stress 0.1903583
## Run 11 stress 0.1986486
## Run 12 stress 0.1924274
## Run 13 stress 0.1937887
## Run 14 stress 0.1927817
## Run 15 stress 0.2300822
## Run 16 stress 0.1929299
## Run 17 stress 0.1880774
## Run 18 stress 0.1927839
## Run 19 stress 0.1929304
## Run 20 stress 0.1865915
## ... New best solution
## ... Procrustes: rmse 0.003521058  max resid 0.02043394
## *** No convergence -- monoMDS stopping criteria:
##      20: stress ratio > sratmax
```

```
plot_ordination(physeq_nifH_ord, phynifH_ord_NMDS)
```



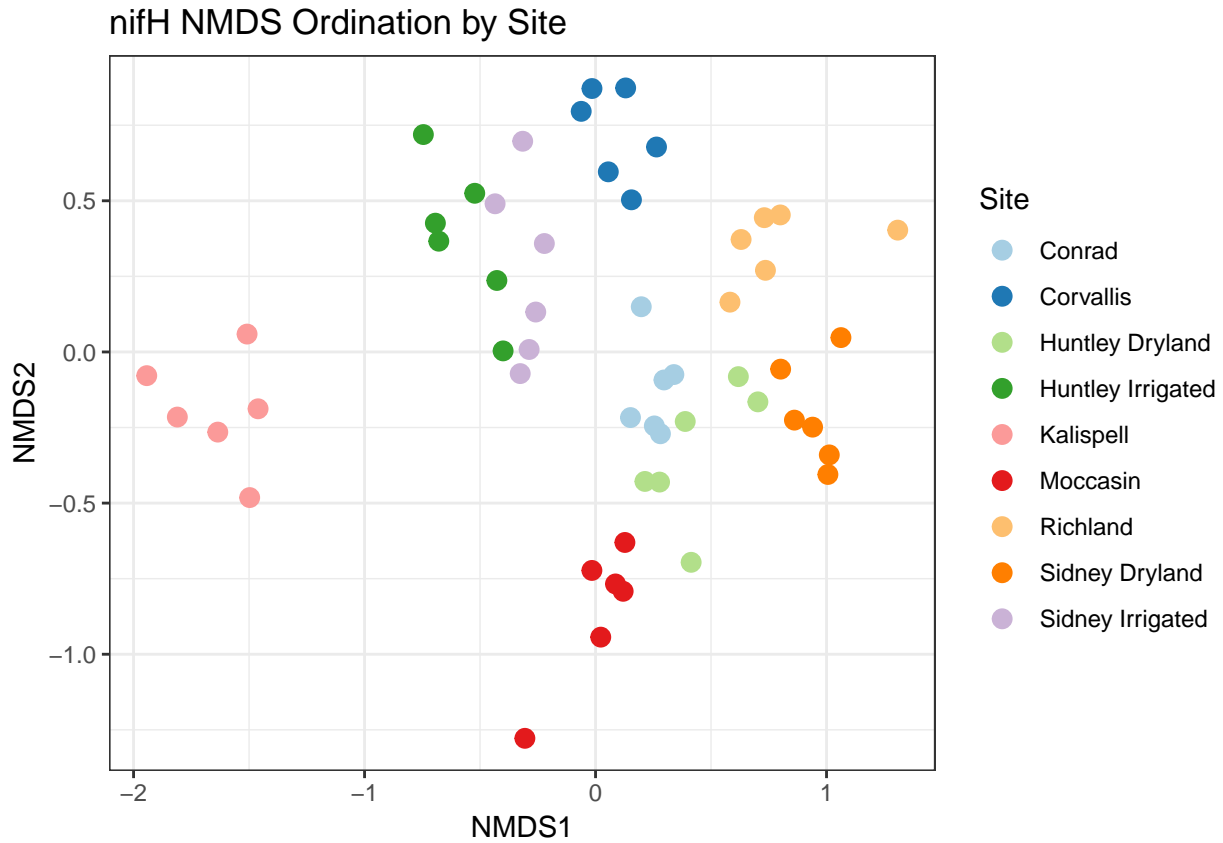
```
phynifH_ord_NMDS
```

```
##
## Call:
## metaMDS(comm = veganifyOTU(physeq), distance = distance)
##
## global Multidimensional Scaling using monoMDS
##
## Data:      wisconsin(sqrt(veganifyOTU(physeq)))
## Distance: bray
##
## Dimensions: 2
## Stress:    0.1865915
## Stress type 1, weak ties
## No convergent solutions - best solution after 20 tries
## Scaling: centring, PC rotation, halfchange scaling
## Species: expanded scores based on 'wisconsin(sqrt(veganifyOTU(physeq)))'
```

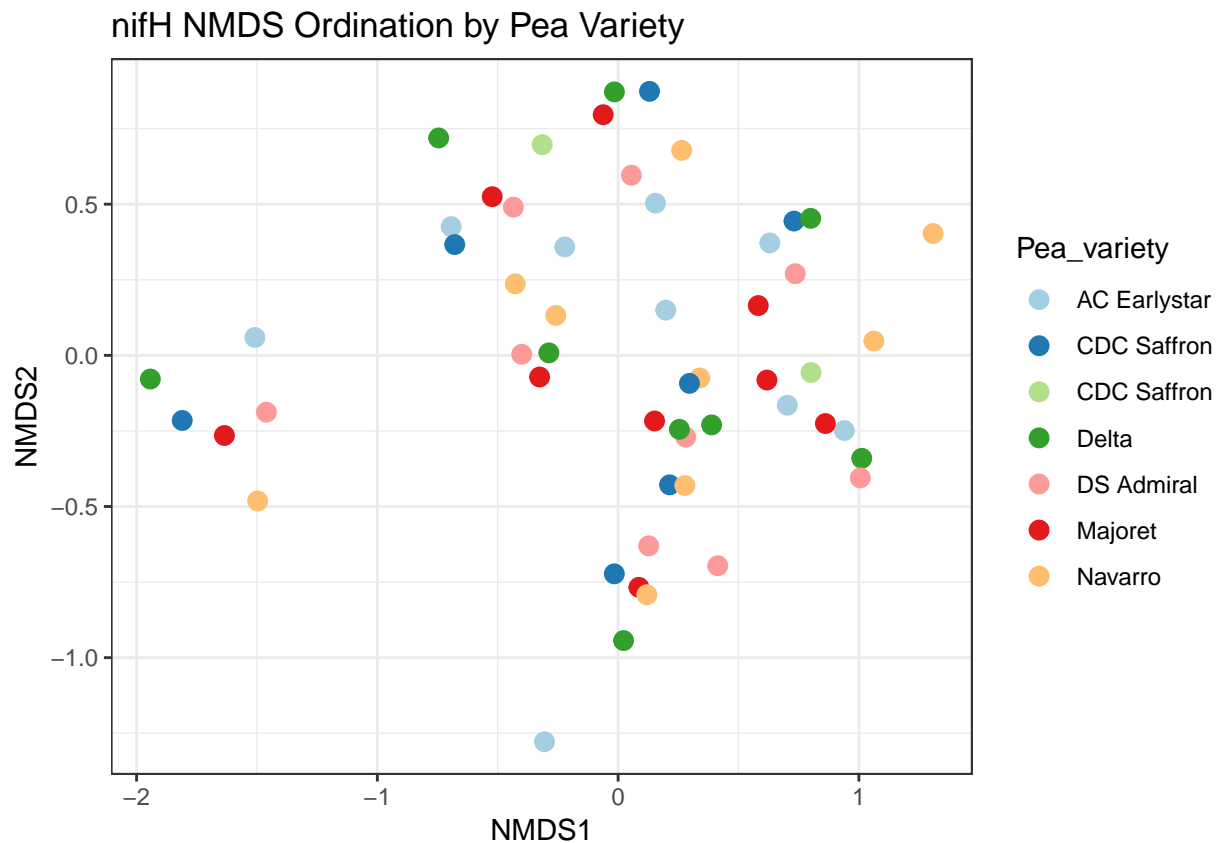
After stress test run, we get a value of 0.16 which is considered good, anything below 0.2 is acceptable.

Plot NMDS with Site

```
plot_ordination(physeq_nifH_ord, phynifH_ord_NMDS, color = "Site")+
  geom_point(size = 3)+
  scale_color_manual(values = farm_col_paired)+
  ggtitle("nifH NMDS Ordination by Site")+
  theme_bw()
```

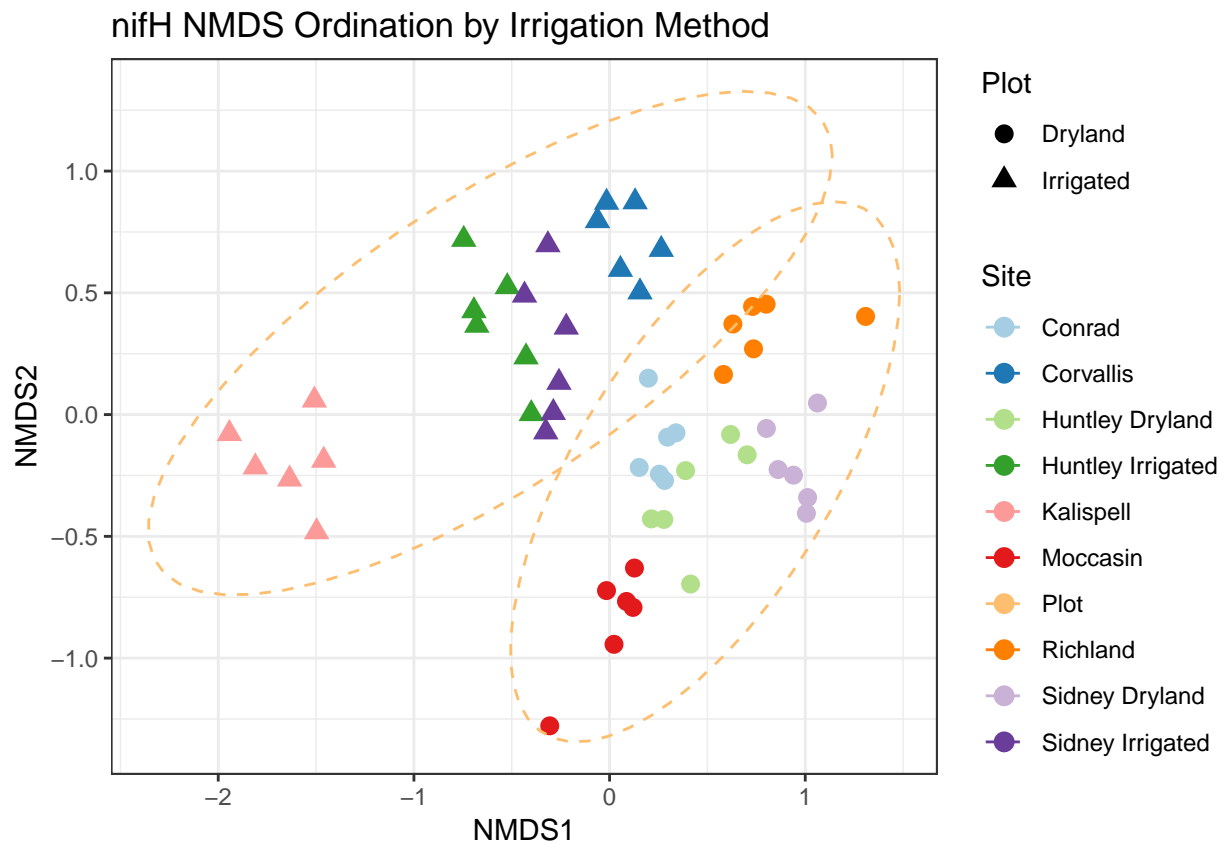


```
plot_ordination(physeq_nifH_ord, phynifH_ord_NMDS, color = "Pea_variety")+
  geom_point(size = 3)+
  scale_color_manual(values = farm_col_paired)+
  ggtitle("nifH NMDS Ordination by Pea Variety")+
  theme_bw()
```



Pea variety does not influence the microbiome composition in NMDS as well

```
plot_ordination(physeq_nifH_ord, phynifH_ord_NMDS, shape = "Plot", color = "Site")+
  geom_point(size = 3)+
  scale_color_manual(values = farm_col_paired)+
  ggtitle("nifH NMDS Ordination by Irrigation Method")+
  stat_ellipse(type = "norm", linetype = 2, aes(color = "Plot")) +
  theme_bw()
```



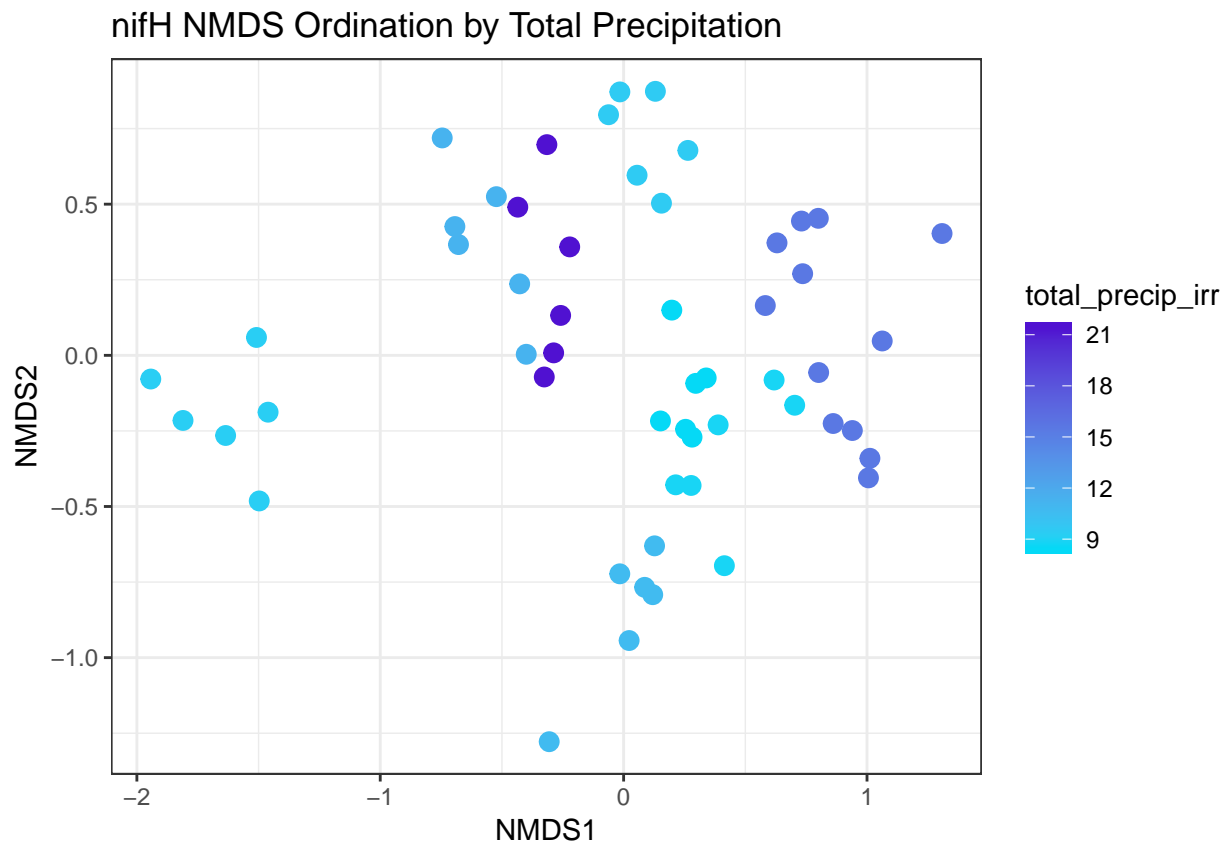
Publish to tiff

```
## pdf
## 2
```

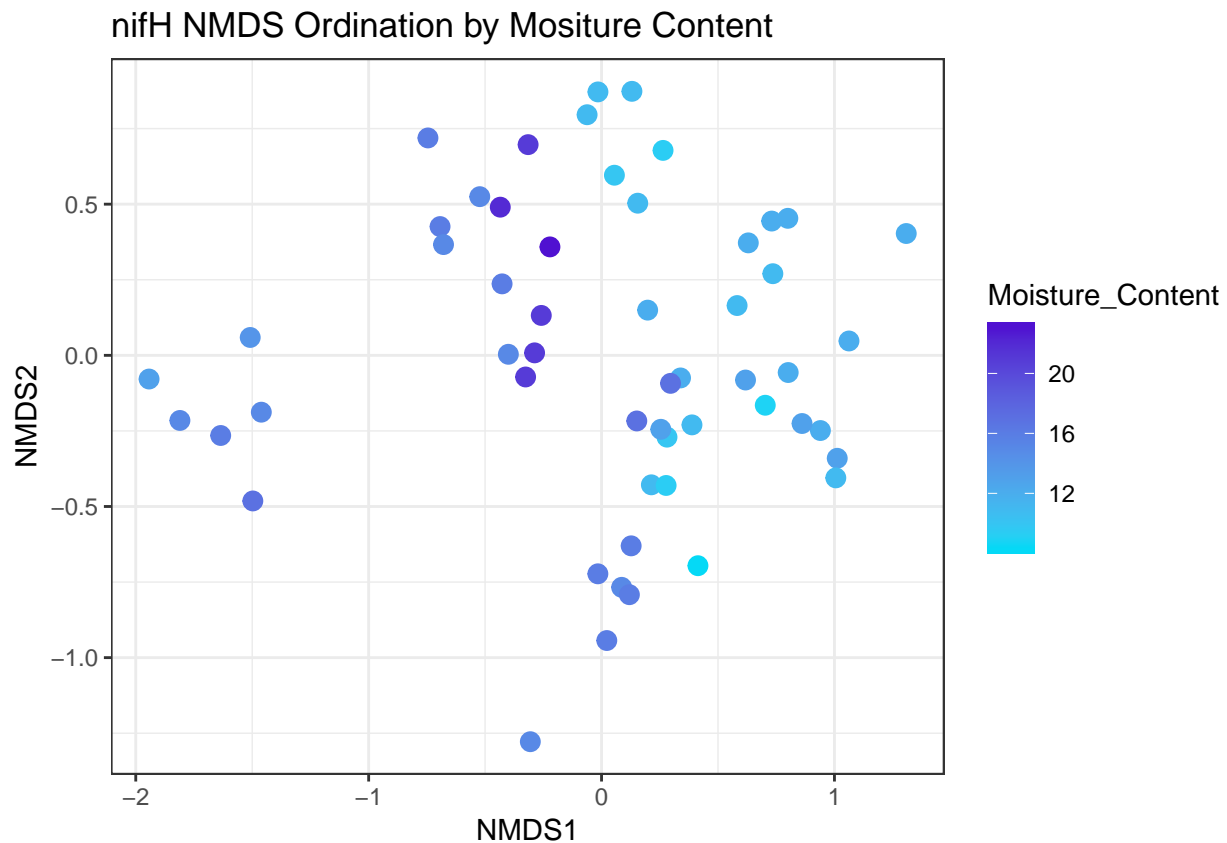
There is again a nice splitting pattern between irrigated and dryland

Total water added

```
plot_ordination(physeq_nifH_ord, phynifH_ord_NMDS, color = "total_precip_irr")+
  geom_point(size = 3)+
  scale_color_gradient(low='#05D9F6', high='#5011D1')+
  ggtitle("nifH NMDS Ordination by Total Precipitation")+
  theme_bw()
```

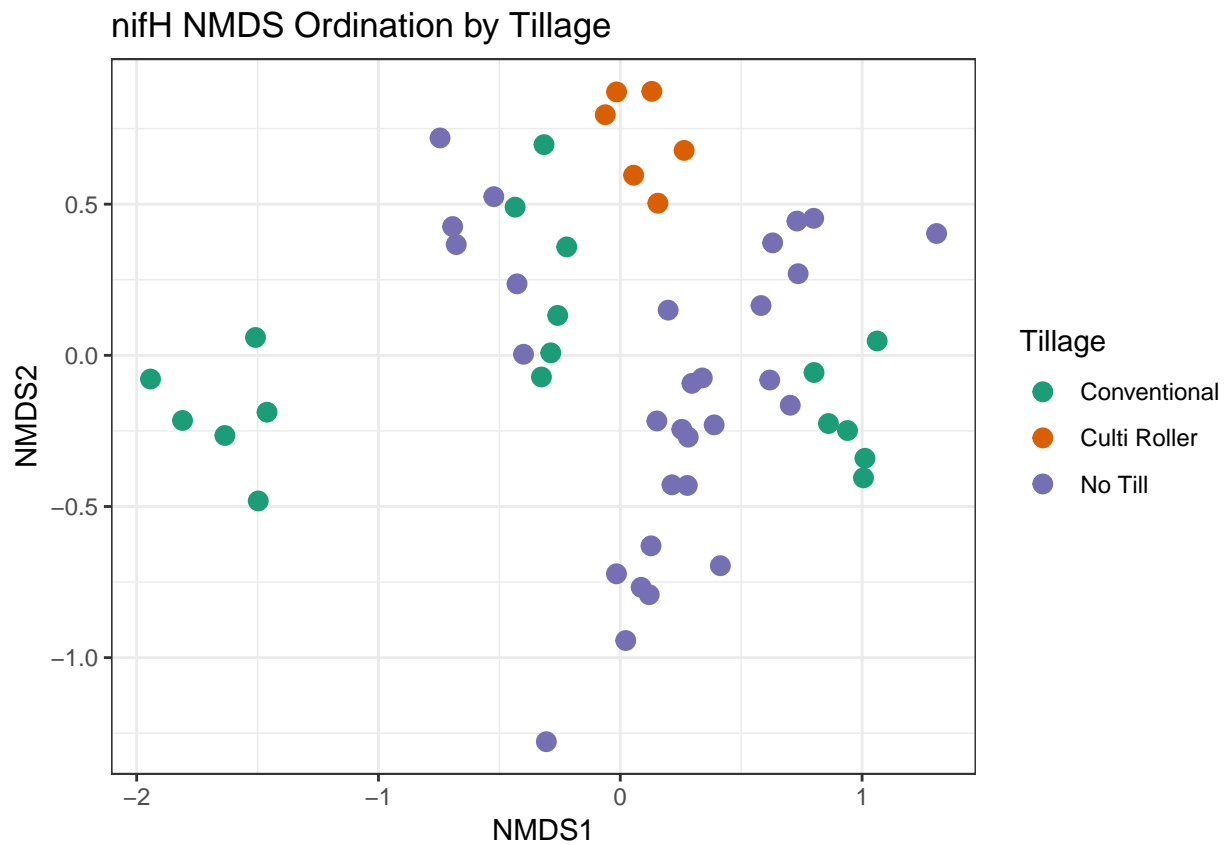
```
plot_ordination(physeq_nifH_ord, phynifH_ord_NMDS, color = "Moisture_Content")+
  geom_point(size = 3)+
  scale_color_continuous(low='#05D9F6', high='#5011D1')+
  ggtitle("nifH NMDS Ordination by Moisture Content")+
  theme_bw()
```



Total water does not ordinate well in NMDS as compared to DCA.

Lets try some of the other variables like till, fallow, and previous crop.

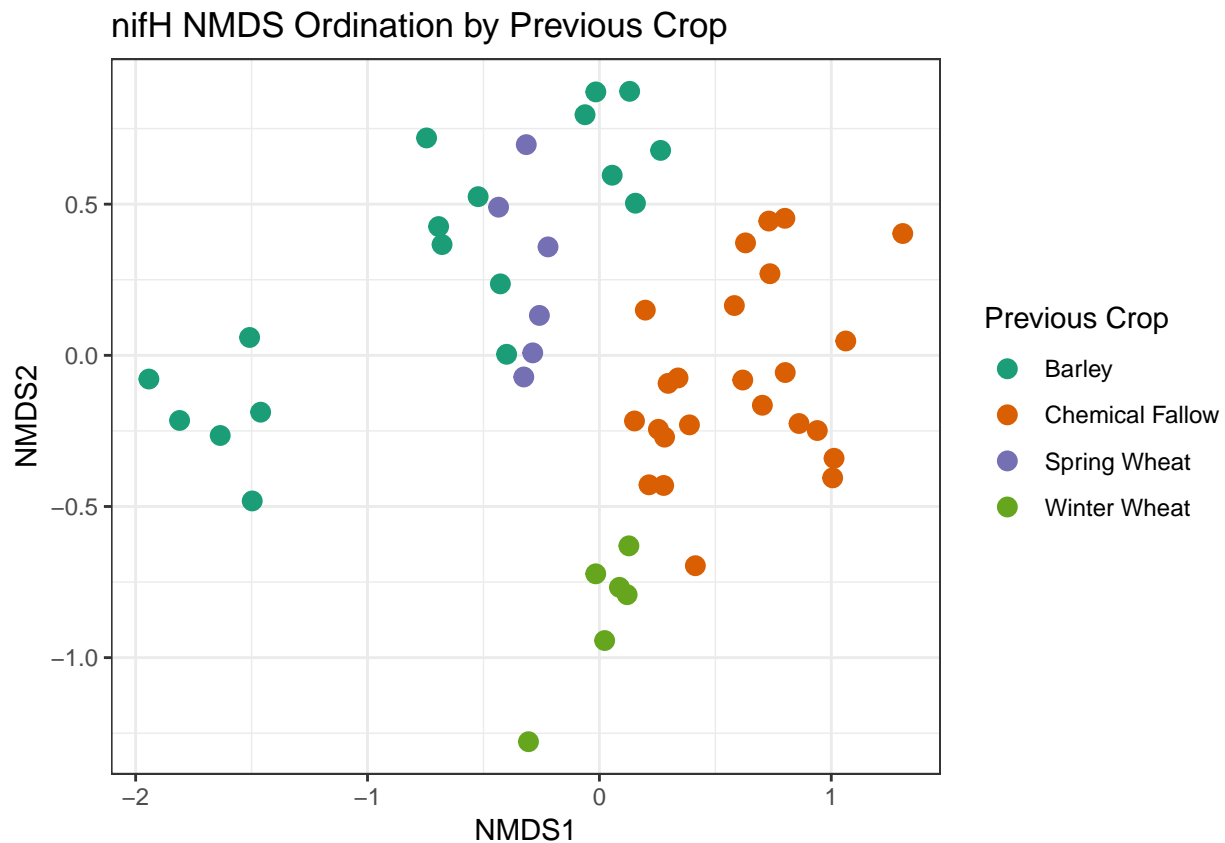
```
plot_ordination(physeq_nifH_ord, phynifH_ord_NMDS,color = "Tillage")+
  geom_point(size = 3)+
  scale_color_manual(values = farm_col_dark, breaks=c("Conventional", "Culti-roller", "No_till"),
    labels=c("Conventional", "Culti Roller", "No Till"))+
  ggtitle("nifH NMDS Ordination by Tillage")+
  theme_bw()
```



Publish to .tiff

```
## pdf
## 2
```

```
plot_ordination(physeq_nifH_ord, phynifH_ord_NMDS, color = "prev_crop")+
  geom_point(size = 3)+
  scale_color_manual(values = c("#1B9E77", "#D95F02", "#7570B3", "#66a61E"),
    name = "Previous Crop",
    breaks=c("barley", "Chem_fallow", "Spring_wheat", "winter_wheat"),
    labels=c("Barley", "Chemical Fallow", "Spring Wheat", "Winter Wheat"))+
  ggtitle("nifH NMDS Ordination by Previous Crop")+
  theme_bw()
```



Publish to .tiff

```
## pdf
## 2
```

Beta dispersions

Test the differences in group homogeneities. Do our farm management factors effect the homogeneity of the bray curtis distance?

If a group (Site) in the MDS space are close but have different dispersion you could have a significant results when it is only a difference in dispersion.

Anderson (2006)-<https://www.ncbi.nlm.nih.gov/pubmed/16542252>

<https://onlinelibrary.wiley.com/doi/epdf/10.1111/j.1461-0248.2006.00926.x>

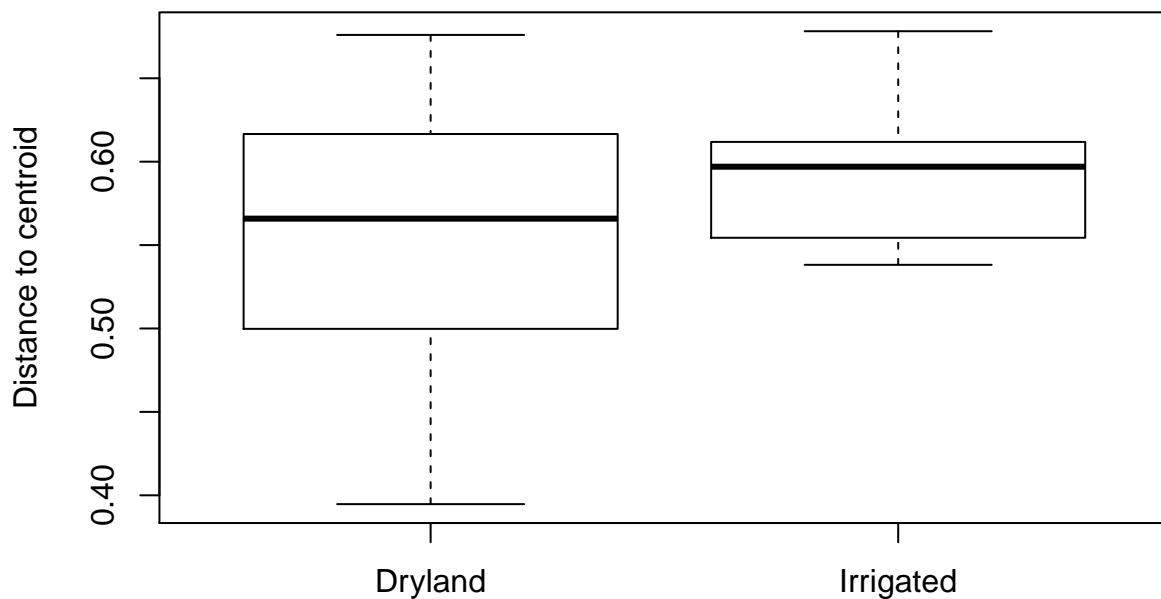
Irrigation beta dispersion

```
nifH_bray<-distance(physeq_nifH_ord, method = "bray")
disp.plot <- betadisper(nifH_bray, meta2$Plot)
permutest(disp.plot, pairwise=TRUE, permutations=1000)
```

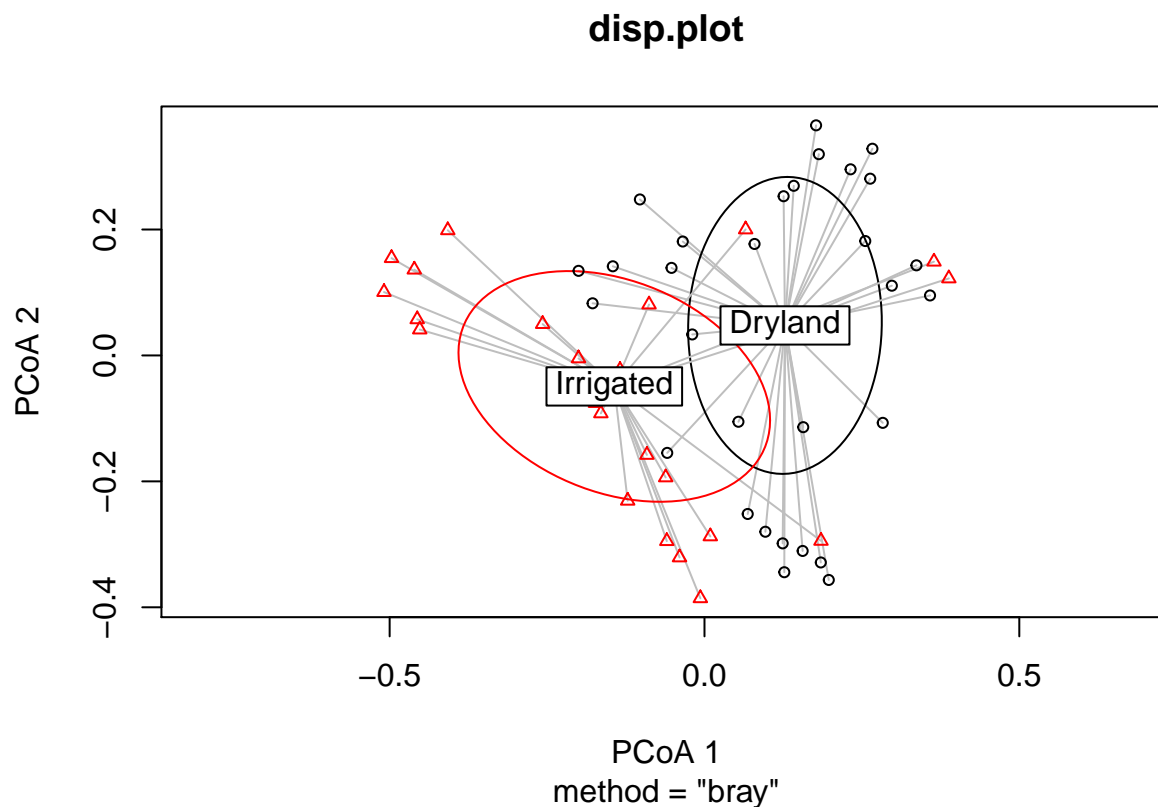
```
##
## Permutation test for homogeneity of multivariate dispersions
```

```
## Permutation: free
## Number of permutations: 1000
##
## Response: Distances
##      Df Sum Sq Mean Sq    F N.Perm Pr(>F)
## Groups  1 0.01565 0.015651 4.0587   1000 0.04396 *
## Residuals 52 0.20051 0.003856
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Pairwise comparisons:
## (Observed p-value below diagonal, permuted p-value above diagonal)
##      Dryland Irrigated
## Dryland              0.049
## Irrigated 0.049129
```

```
boxplot(displot)
```



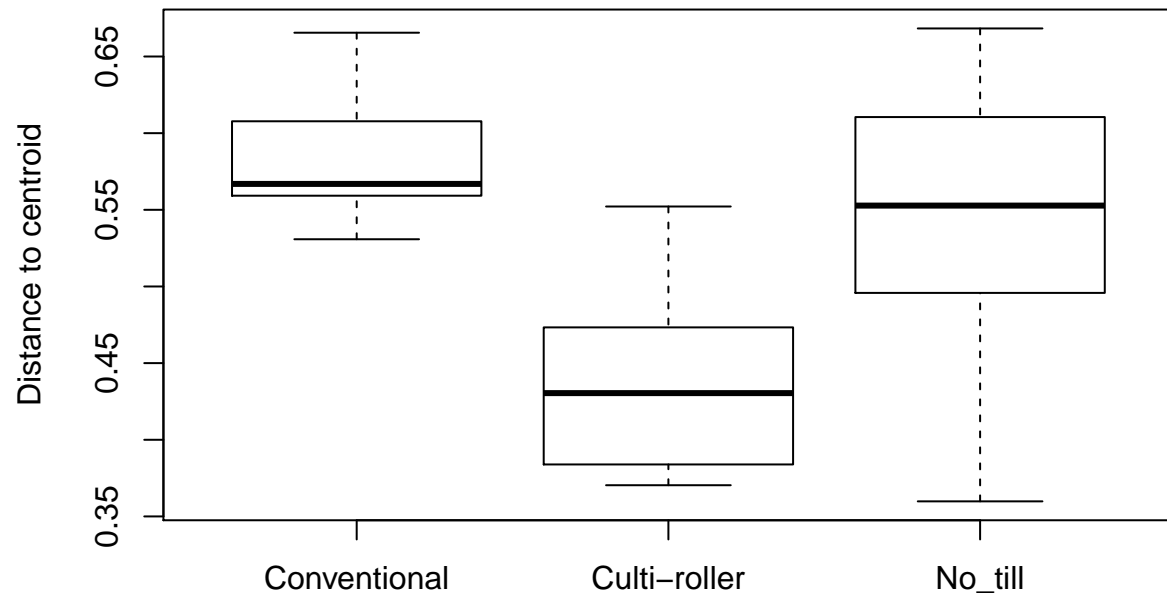
```
plot(displot, hull = FALSE, ellipse = TRUE)
```



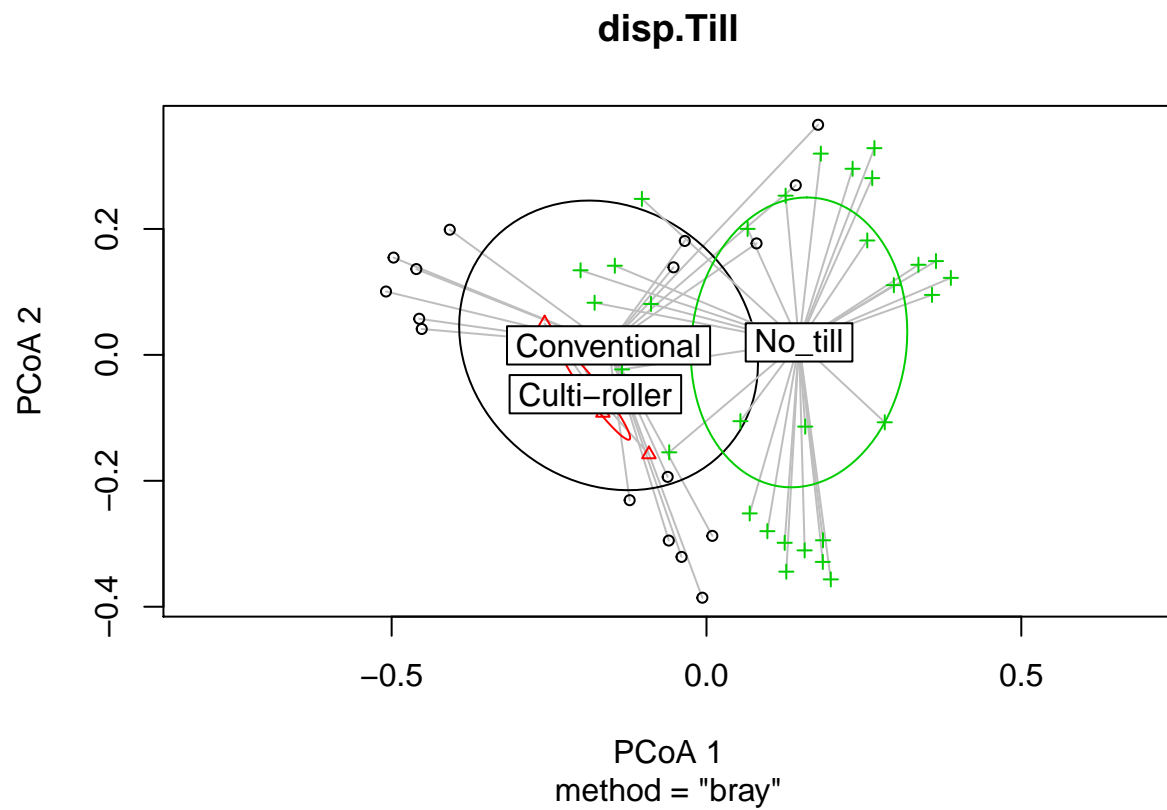
```
disp.Till <- betadisper(distance(physeq_nifH_ord, method = "bray"), meta2$Tillage)
permutest(disp.Till, pairwise=TRUE, permutations=1000)
```

```
##
## Permutation test for homogeneity of multivariate dispersions
## Permutation: free
## Number of permutations: 1000
##
## Response: Distances
##      Df  Sum Sq Mean Sq    F N.Perm  Pr(>F)
## Groups   2 0.089137 0.044568 8.6297   1000 0.001998 **
## Residuals 51 0.263390 0.005165
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Pairwise comparisons:
## (Observed p-value below diagonal, permuted p-value above diagonal)
##      Conventional Culti-roller No_till
## Conventional          9.9900e-04 0.0669
## Culti-roller    2.0614e-06          0.0050
## No_till          6.7763e-02  1.1376e-02
```

```
boxplot(displacement.Till)
```



```
plot(displacement.Till, hull = FALSE, ellipse = TRUE)
```

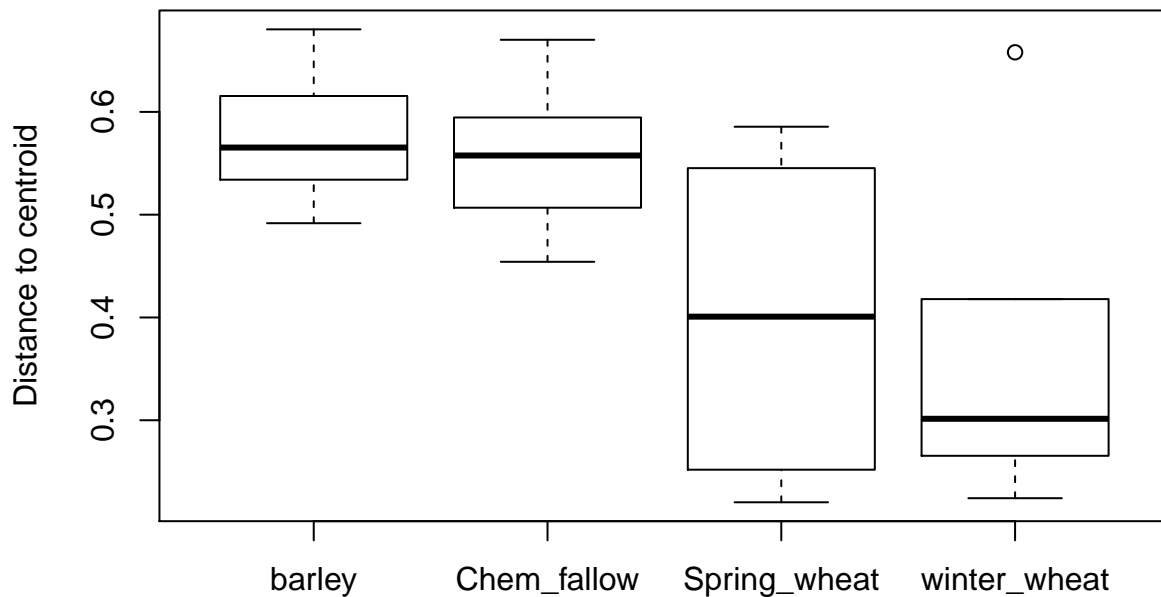


Tillage (Conventional and Culti-Roller) vs NO till are splitting

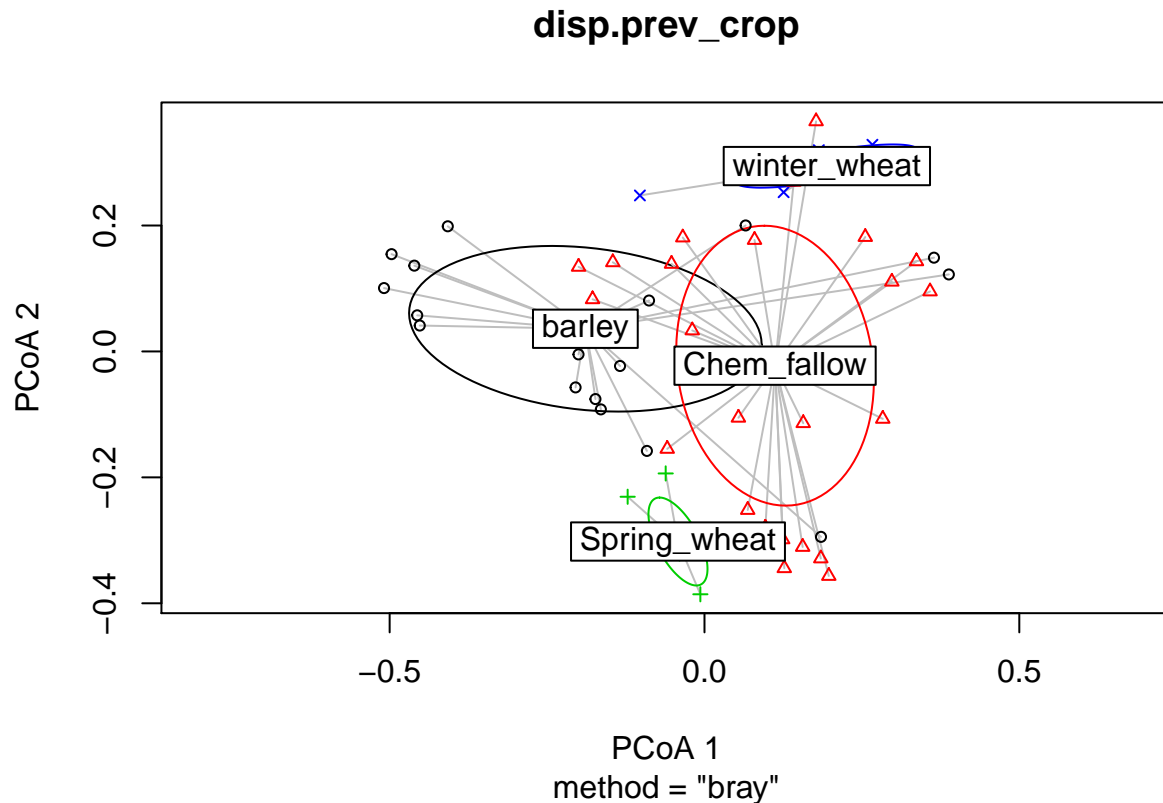
```
disp.prev_crop <- betadisper(distance(physeq_nifH_ord, method = "bray"), meta2$prev_crop)
permutest(disp.prev_crop, pairwise=TRUE, permutations=1000)
```

```
##
## Permutation test for homogeneity of multivariate dispersions
## Permutation: free
## Number of permutations: 1000
##
## Response: Distances
##           Df Sum Sq Mean Sq      F N.Perm  Pr(>F)
## Groups      3 0.32146 0.107153 13.645   1000 0.000999 ***
## Residuals  50 0.39265 0.007853
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Pairwise comparisons:
## (Observed p-value below diagonal, permuted p-value above diagonal)
##           barley Chem_fallow Spring_wheat winter_wheat
## barley           1.9580e-01  9.9900e-04    0.0010
## Chem_fallow  1.8314e-01           2.9970e-03    0.0020
## Spring_wheat 3.6221e-04  8.3887e-04           0.7073
## winter_wheat 3.1348e-05  5.3897e-05  6.7995e-01
```

```
boxplot(disp.prev_crop)
```



```
plot(disp.prev_crop, hull = FALSE, ellipse = TRUE)
```

```
TukeyHSD(disp.prev_crop)
```

```
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = distances ~ group, data = df)
##
## $group
```

	diff	lwr	upr	p adj
Chem_fallow-barley	-0.02453687	-0.09796924	0.04889550	0.8110677
Spring_wheat-barley	-0.17676352	-0.28778283	-0.06574422	0.0005585
winter_wheat-barley	-0.21613604	-0.32715535	-0.10511673	0.0000237
Spring_wheat-Chem_fallow	-0.15222665	-0.25972063	-0.04473267	0.0024172
winter_wheat-Chem_fallow	-0.19159917	-0.29909315	-0.08410518	0.0001054
winter_wheat-Spring_wheat	-0.03937252	-0.17534285	0.09659781	0.8677798

```
disp.site.nifH <- betadisper(distance(physeq_nifH_ord, method = "bray"), meta2$Site)
permutest(disp.site.nifH, pairwise=TRUE, permutations=1000)
```

```
##
## Permutation test for homogeneity of multivariate dispersions
## Permutation: free
## Number of permutations: 1000
##
## Response: Distances
```

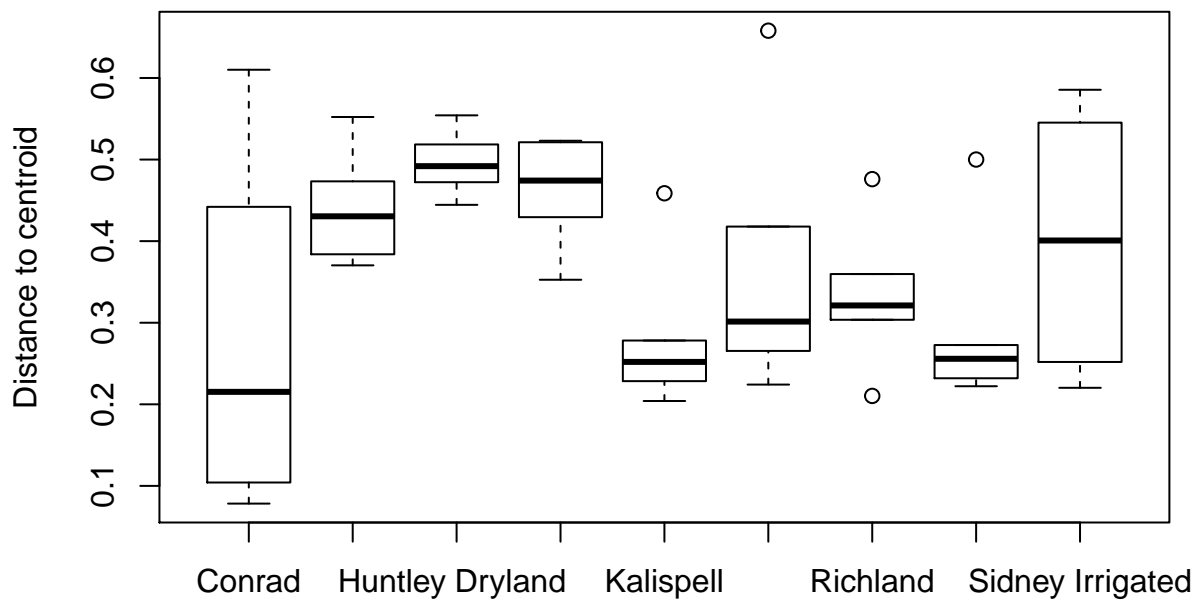
	Df	Sum Sq	Mean Sq	F	N.Perm	Pr(>F)
Groups	8	0.32992	0.041240	2.8058	1000	0.01598 *

```

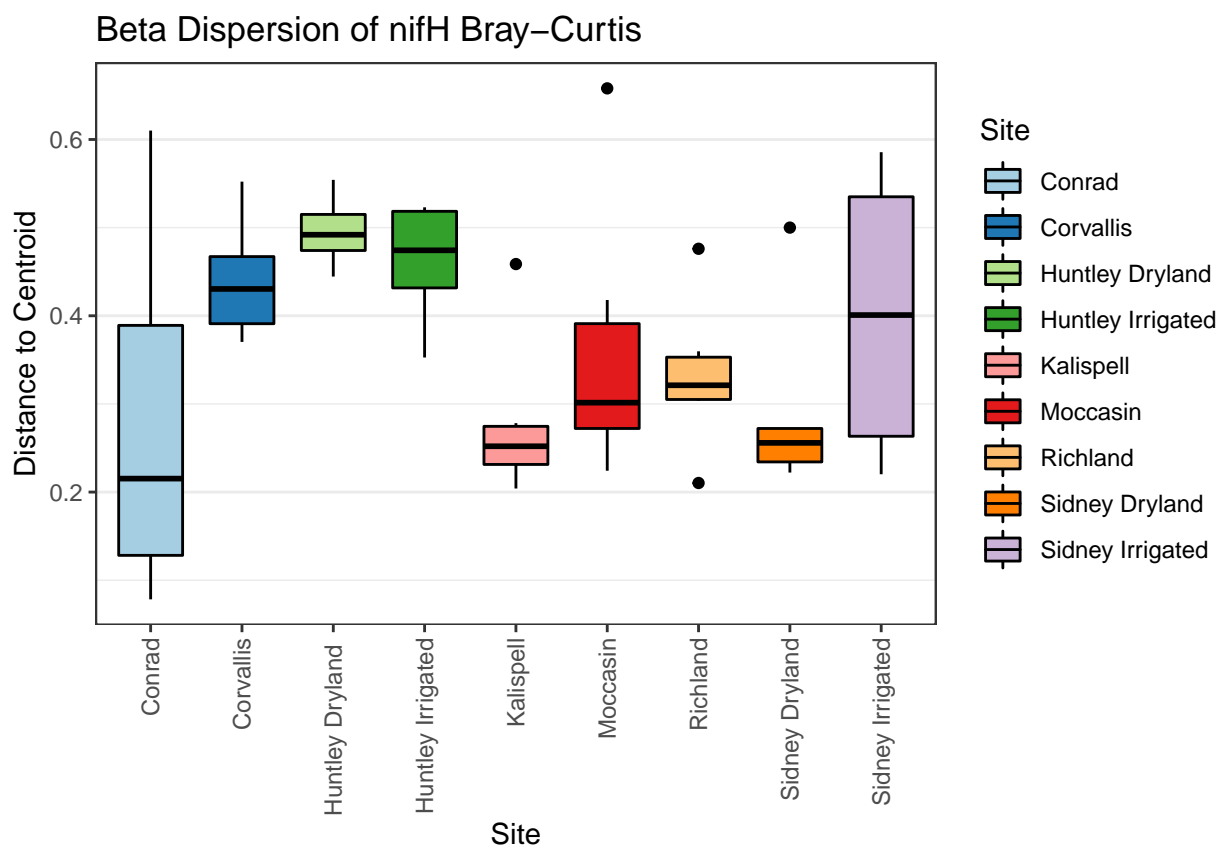
## Residuals 45 0.66143 0.014698
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Pairwise comparisons:
## (Observed p-value below diagonal, permuted p-value above diagonal)
##
##           Conrad  Corvallis Huntley Dryland  Huntley Irrigated
## Conrad           0.08691309      0.03096903      0.05694306
## Corvallis        0.09803153      0.11988012      0.61138861
## Huntley Dryland  0.02995297 0.10991872      0.32667333
## Huntley Irrigated 0.06494609 0.57900203      0.32466376
## Kalispell        0.98867876 0.00601294      0.00033484      0.00280268
## Moccasin         0.45050342 0.29029786      0.07247948      0.18292552
## Richland         0.56654632 0.03642784      0.00177420      0.01586284
## Sidney Dryland   0.90035851 0.01441515      0.00113378      0.00698916
## Sidney Irrigated 0.27835921 0.59461068      0.19328302      0.40961155
##
##           Kalispell  Moccasin  Richland Sidney Dryland
## Conrad           0.98801199 0.46253746 0.55644356      0.89310689
## Corvallis        0.00699301 0.27372627 0.03796204      0.01998002
## Huntley Dryland  0.00099900 0.07692308 0.00499500      0.00299700
## Huntley Irrigated 0.00799201 0.20279720 0.01898102      0.00899101
## Kalispell        0.28271728 0.32767233 0.85714286
## Moccasin         0.29701845      0.67632368 0.35464535
## Richland         0.32751945 0.69935339      0.43556444
## Sidney Dryland   0.85273083 0.37883278 0.46494154
## Sidney Irrigated 0.13987695 0.67994777 0.38070532      0.18913329
##
##           Sidney Irrigated
## Conrad           0.2697
## Corvallis        0.5674
## Huntley Dryland  0.1928
## Huntley Irrigated 0.4176
## Kalispell        0.1349
## Moccasin         0.6643
## Richland         0.3666
## Sidney Dryland   0.2008
## Sidney Irrigated

```

```
boxplot(dispatch.nifH)
```



```
nifH_dispersion_site<-data.frame(Distance_to_centroid=disp.site.nifH$distances, Site=disp.site.nifH$grove)
ggboxplot(nifH_dispersion_site, x = "Site", y = "Distance_to_centroid",
  rug = TRUE,
  fill = "Site", ylab = "Distance to Centroid", title = "Beta Dispersion of nifH Bray-Curtis",
  palette = farm_col_paired)+
theme_bw()+
theme( panel.grid.major.x = element_blank())+
rotate_x_text()
```

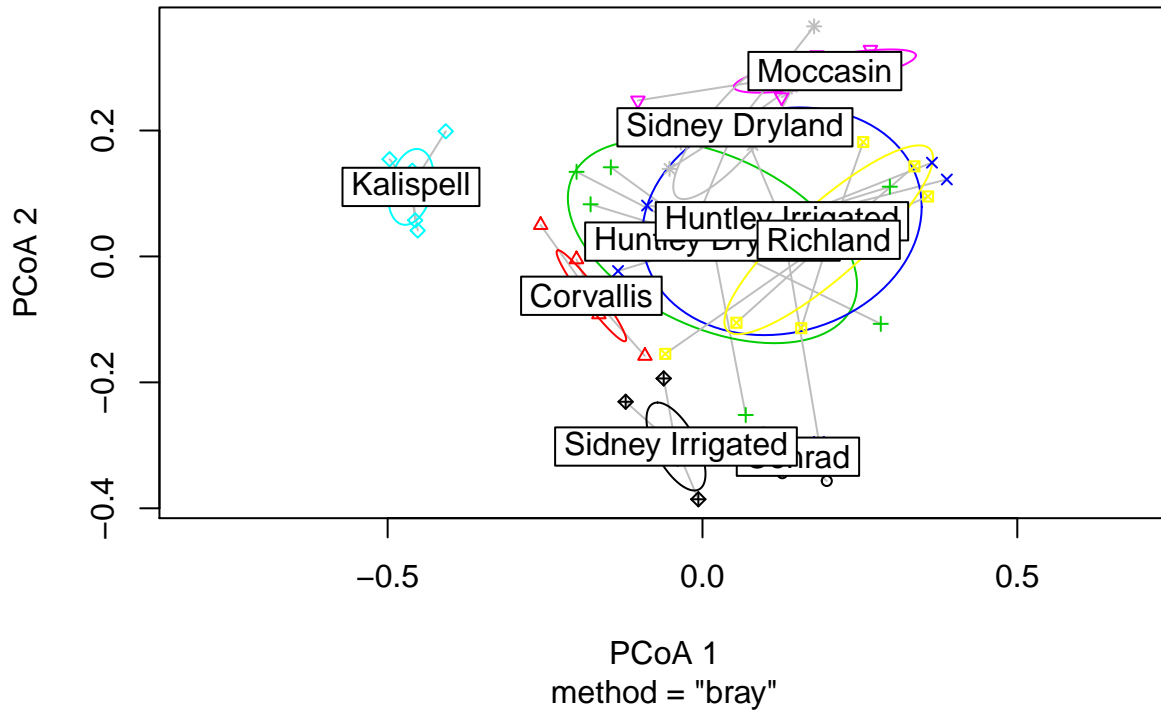


Publish to .tiff

```
## pdf
## 2
```

```
plot(dispen.site.nifH, hull = FALSE, ellipse = TRUE)
```

disp.site.nifH



TukeyHSD(disp.site.nifH)

```
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = distances ~ group, data = df)
##
## $group
```

	diff	lwr	upr
## Corvallis-Conrad	0.162574317	-0.065413487	0.39056212
## Huntley Dryland -Conrad	0.218030908	-0.009956896	0.44601871
## Huntley Irrigated-Conrad	0.184950208	-0.043037596	0.41293801
## Kalispell-Conrad	0.001348719	-0.226639086	0.22933652
## Moccasin-Conrad	0.083865359	-0.144122446	0.31185316
## Richland-Conrad	0.054458523	-0.173529281	0.28244633
## Sidney Dryland-Conrad	0.012201487	-0.215786317	0.24018929
## Sidney Irrigated-Conrad	0.123237876	-0.104749928	0.35122568
## Huntley Dryland -Corvallis	0.055456591	-0.172531214	0.28344440
## Huntley Irrigated-Corvallis	0.022375890	-0.205611914	0.25036369
## Kalispell-Corvallis	-0.161225599	-0.389213403	0.06676221
## Moccasin-Corvallis	-0.078708959	-0.306696763	0.14927885
## Richland-Corvallis	-0.108115794	-0.336103598	0.11987201
## Sidney Dryland-Corvallis	-0.150372830	-0.378360635	0.07761497
## Sidney Irrigated-Corvallis	-0.039336441	-0.267324246	0.18865136
## Huntley Irrigated-Huntley Dryland	-0.033080700	-0.261068505	0.19490710
## Kalispell-Huntley Dryland	-0.216682189	-0.444669994	0.01130561
## Moccasin-Huntley Dryland	-0.134165550	-0.362153354	0.09382225

## Richland-Huntley Dryland	-0.163572385	-0.391560189	0.06441542
## Sidney Dryland-Huntley Dryland	-0.205829421	-0.433817225	0.02215838
## Sidney Irrigated-Huntley Dryland	-0.094793032	-0.322780836	0.13319477
## Kalispell-Huntley Irrigated	-0.183601489	-0.411589293	0.04438632
## Moccasin-Huntley Irrigated	-0.101084849	-0.329072654	0.12690296
## Richland-Huntley Irrigated	-0.130491684	-0.358479489	0.09749612
## Sidney Dryland-Huntley Irrigated	-0.172748721	-0.400736525	0.05523908
## Sidney Irrigated-Huntley Irrigated	-0.061712332	-0.289700136	0.16627547
## Moccasin-Kalispell	0.082516640	-0.145471164	0.31050444
## Richland-Kalispell	0.053109805	-0.174878000	0.28109761
## Sidney Dryland-Kalispell	0.010852768	-0.217135036	0.23884057
## Sidney Irrigated-Kalispell	0.121889157	-0.106098647	0.34987696
## Richland-Moccasin	-0.029406835	-0.257394639	0.19858097
## Sidney Dryland-Moccasin	-0.071663871	-0.299651676	0.15632393
## Sidney Irrigated-Moccasin	0.039372517	-0.188615287	0.26736032
## Sidney Dryland-Richland	-0.042257036	-0.270244841	0.18573077
## Sidney Irrigated-Richland	0.068779353	-0.159208452	0.29676716
## Sidney Irrigated-Sidney Dryland	0.111036389	-0.116951415	0.33902419
##	p adj		
## Corvallis-Conrad	0.3508722		
## Huntley Dryland -Conrad	0.0707667		
## Huntley Irrigated-Conrad	0.1984083		
## Kalispell-Conrad	1.0000000		
## Moccasin-Conrad	0.9527776		
## Richland-Conrad	0.9969598		
## Sidney Dryland-Conrad	1.0000000		
## Sidney Irrigated-Conrad	0.7065213		
## Huntley Dryland -Corvallis	0.9965542		
## Huntley Irrigated-Corvallis	0.9999961		
## Kalispell-Corvallis	0.3617848		
## Moccasin-Corvallis	0.9672588		
## Richland-Corvallis	0.8282749		
## Sidney Dryland-Corvallis	0.4553083		
## Sidney Irrigated-Corvallis	0.9997068		
## Huntley Irrigated-Huntley Dryland	0.9999200		
## Kalispell-Huntley Dryland	0.0740866		
## Moccasin-Huntley Dryland	0.6062392		
## Richland-Huntley Dryland	0.3429114		
## Sidney Dryland-Huntley Dryland	0.1059547		
## Sidney Irrigated-Huntley Dryland	0.9086105		
## Kalispell-Huntley Irrigated	0.2059932		
## Moccasin-Huntley Irrigated	0.8742599		
## Richland-Huntley Irrigated	0.6405710		
## Sidney Dryland-Huntley Irrigated	0.2746068		
## Sidney Irrigated-Huntley Irrigated	0.9929035		
## Moccasin-Kalispell	0.9569325		
## Richland-Kalispell	0.9974455		
## Sidney Dryland-Kalispell	1.0000000		
## Sidney Irrigated-Kalispell	0.7183816		
## Richland-Moccasin	0.9999674		
## Sidney Dryland-Moccasin	0.9814416		
## Sidney Irrigated-Moccasin	0.9997048		
## Sidney Dryland-Richland	0.9995028		
## Sidney Irrigated-Richland	0.9856585		

```
## Sidney Irrigated-Sidney Dryland    0.8069758
```

The results from the beta dispersion show that we have a significant difference in the heterogeneity of our sites due to each of the farm management factors. We see that there are sites that do not have significant differences and others that do. Beta dispersion cannot except models so we cannot detect if the differences are due to nestedness (prev_crop/Site).

Due to the design of the experiment it will be hard to determine if the farm management practices are responsible for the variation in bacterial population.

PERMANOVA (adonis)

We will test the significance of the farm management using the permuted-multivariate-ANOVA function in vegan called adonis. Adonis can test models and nestedness though the above beta dispersion test showed that most of the significance is due to the dispersion we cannot say much. But we can show we have the location effect.

```
adonis(distance(physeq_nifH_ord, method = "bray")
~Plot*prev_crop*Tillage*Pea_variety, data = meta2, permutations = 1000)
```

```
##
## Call:
## adonis(formula = distance(physeq_nifH_ord, method = "bray") ~      Plot * prev_crop * Tillage * Pea_
##
## Permutation: free
## Number of permutations: 1000
##
## Terms added sequentially (first to last)
##
##              Df SumsOfSqs MeanSqs F.Model      R2    Pr(>F)
## Plot              1      1.4093 1.40932   3.9327 0.07299 0.000999 ***
## prev_crop          2      2.4983 1.24917   3.4858 0.12938 0.000999 ***
## Tillage            2      3.1325 1.56624   4.3706 0.16222 0.000999 ***
## Pea_variety        6      1.3105 0.21841   0.6095 0.06787 1.000000
## Plot:Tillage        1      1.5306 1.53056   4.2711 0.07926 0.000999 ***
## Plot:Pea_variety    6      0.8231 0.13718   0.3828 0.04263 1.000000
## prev_crop:Pea_variety 9      1.6820 0.18689   0.5215 0.08711 1.000000
## Tillage:Pea_variety 10     1.9518 0.19518   0.5446 0.10108 1.000000
## Plot:Tillage:Pea_variety 4     0.6713 0.16782   0.4683 0.03476 1.000000
## Residuals          12     4.3003 0.35836           0.22270
## Total              53    19.3096           1.00000
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
adonis(distance(physeq_nifH_ord, method = "bray")
~(Plot*prev_crop*Tillage*Pea_variety)/Site, data = meta2, permutations = 1000)
```

```
##
```

```
## Call:
## adonis(formula = distance(physeq_nifH_ord, method = "bray") ~ (Plot * prev_crop * Tillage * Pea,
##
## Permutation: free
## Number of permutations: 1000
##
## Terms added sequentially (first to last)
##
##
```

	Df	SumsOfSqs	MeanSqs	F.Model
## Plot	1	1.4093	1	0
## prev_crop	2	2.4983	1	0
## Tillage	2	3.1325	2	0
## Pea_variety	6	1.3105	0	0
## Plot:Tillage	1	1.5306	2	0
## Plot:Pea_variety	6	0.8231	0	0
## prev_crop:Pea_variety	9	1.6820	0	0
## Tillage:Pea_variety	10	1.9518	0	0
## Plot:Tillage:Pea_variety	4	0.6713	0	0
## Plot:prev_crop:Tillage:Pea_variety:Site	12	4.3003	0	0
## Residuals	0	0.0000	-Inf	
## Total	53	19.3096		

```
##
##
```

	R2	Pr(>F)
## Plot	0.07299	1
## prev_crop	0.12938	1
## Tillage	0.16222	1
## Pea_variety	0.06787	1
## Plot:Tillage	0.07926	1
## Plot:Pea_variety	0.04263	1
## prev_crop:Pea_variety	0.08711	1
## Tillage:Pea_variety	0.10108	1
## Plot:Tillage:Pea_variety	0.03476	1
## Plot:prev_crop:Tillage:Pea_variety:Site	0.22270	1
## Residuals	0.00000	
## Total	1.00000	

Data is nested within Site (Location effect) so the significance in the bray-curtis dissimilarity with respect to plot is not significant with the data nested due to the lack of reproduction of conditions at each plot.

Since there is issues with doing permutated anova over multivariate data lets try to fit the chemical and farm management data the NMDS ordination space using the envfit function in vegan

Model Selection

ENVFIT

Envfit does not like single variable values so we remove them

```
meta3<-meta2[, -c(3,4,10,11,27,29,35,38,42,46)]
```

Will remove Site catagories like elevation, lat, long etc that do not diffrentate between site we can call these all geographical factors as they do not change between sites.

```
meta3<-meta3[, -c(2,11:13)]
```

Model fitting will be biased by chemical outliers that are in some plots the best way to avoid this is to determine the outliers (See chemical_analysis.Rmd) and remove the whole variable since functions ENVFIT and BIOENV will remove it if there are any n/a values.

Removing Sulfate_Sulfur, Boron, Molybdenum, Potassium, Vanadium, Chromium and Sodium (From chemical_analysis.rmd)

```
meta3<-meta3[, -c(16,18,21,29,31)]
```

```
envfitnifH <- envfit(phynifH_ord_NMDS , meta3, na.rm = TRUE, permu= 10000)
envfitnifH
```

```
##
## ***VECTORS
##
##          NMDS1    NMDS2    r2    Pr(>r)
## season_precip  0.43714 -0.89939 0.0400  0.354465
## irrigation     0.30011  0.95391 0.4424 9.999e-05 ***
## total_precip_irr 0.65883  0.75229 0.1486  0.016098 *
## grain_yield    -0.39481  0.91876 0.0227  0.562644
## Organic_Matter -0.70859 -0.70562 0.4791 9.999e-05 ***
## Moisture_Content 0.58097 -0.81393 0.0484  0.287471
## Nitrate_Nitrite -0.02415  0.99971 0.2201  0.001800 **
## Ammonia        -0.04719 -0.99889 0.0214  0.580542
## Av_Phosphorus   0.43085  0.90242 0.4094 9.999e-05 ***
## Av_Potassium    0.76331  0.64603 0.4202 9.999e-05 ***
## pH              0.02901 -0.99958 0.0218  0.582442
## Barium          -0.08073 -0.99674 0.4684 9.999e-05 ***
## Calcium         0.38027 -0.92488 0.0923  0.084192 .
## Cobalt          0.16341 -0.98656 0.0247  0.523448
## Copper          0.54680 -0.83726 0.0748  0.137786
## Iron            0.52535 -0.85088 0.0625  0.185881
## Magnesium       0.21991  0.97552 0.0438  0.321168
## Manganese       0.95991  0.28030 0.0048  0.886211
## Nickel          0.62941 -0.77707 0.1258  0.036296 *
## Phosphorus      0.80483  0.59350 0.0058  0.865413
## Sulfur          0.09856 -0.99513 0.1700  0.009399 **
## Zinc            0.56036 -0.82825 0.2387  0.001200 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```

## Permutation: free
## Number of permutations: 10000
##
## ***FACTORS:
##
## Centroids:
##
##           NMDS1  NMDS2
## SiteConrad      0.9471 -0.2049
## SiteCorvallis    0.0880  0.7196
## SiteHuntley Dryland -0.1353  0.1147
## SiteHuntley Irrigated -0.0068 -0.0739
## SiteKalispell    -1.6425 -0.1949
## SiteMoccasin      0.0057 -0.8556
## SiteRichland      0.2534 -0.1251
## SiteSidney Dryland -0.3073  0.2691
## SiteSidney Irrigated 0.7978  0.3512
## Pea_varietyAC Earlystar -0.1756  0.1583
## Pea_varietyCDC Saffron 0.0279 -0.1372
## Pea_varietyCDC Saffron 0.2077  0.5706
## Pea_varietyDelta   0.1676 -0.0482
## Pea_varietyDS Admiral 0.0456 -0.1395
## Pea_varietyMajoret -0.1055  0.0036
## Pea_varietyNavarro 0.0000  0.0056
## PlotDryland       0.1527 -0.1604
## PlotIrrigated     -0.1909  0.2005
## TillageConventional -0.3840  0.1418
## TillageCulti-roller 0.0880  0.7196
## TillageNo_till     0.2128 -0.2290
## prev_cropbarley    -0.5204  0.1502
## prev_cropChem_fallow 0.1895  0.0134
## prev_cropSpring_wheat 0.7978  0.3512
## prev_cropwinter_wheat 0.0057 -0.8556
##
## Goodness of fit:
##           r2    Pr(>r)
## Site      0.8274 9.999e-05 ***
## Pea_variety 0.0449  0.9671
## Plot      0.0764  0.0223 *
## Tillage   0.2100  0.0002 ***
## prev_crop 0.3484 9.999e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Permutation: free
## Number of permutations: 10000

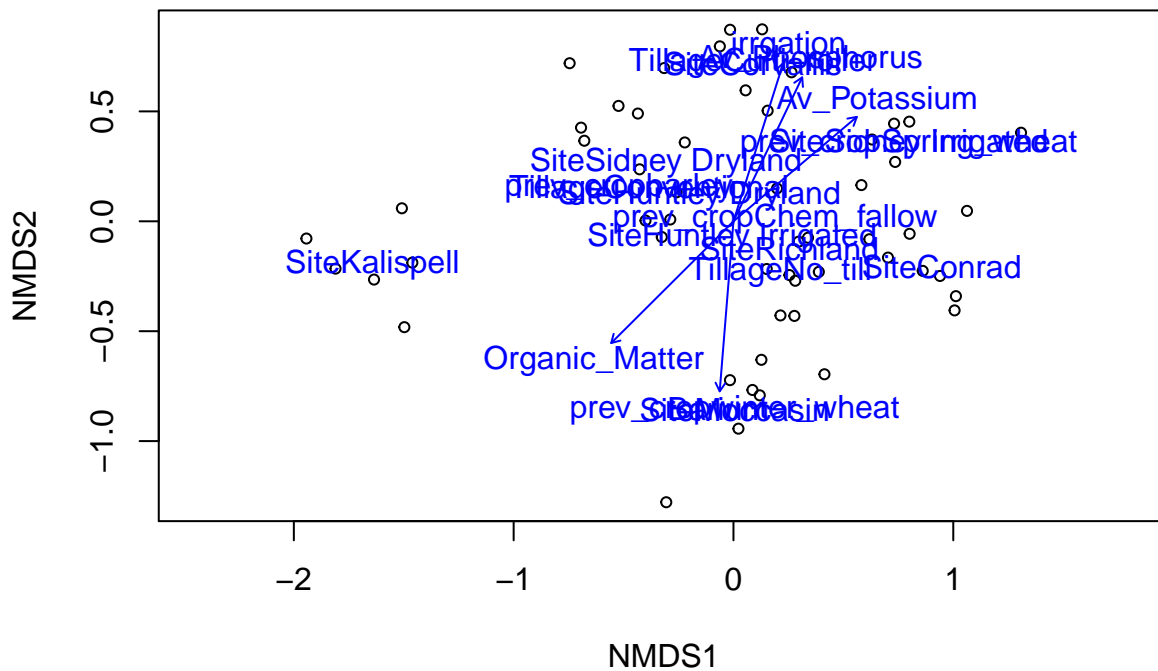
```

write to table

The envfit function allows us to see the correlation of our environmental vectors to the bray-curtis species dissimilarity matrix in NMDS space. This is a loose correlation to real linear correlation but it can tell us how the NMDS ordination is being driven.

Try a quick plot with base r and vegan for the vectors

```
plot(phynifH_ord_NMDS, display = "sites")
plot(envfitnifH, p.max = 0.001 )
```



Organic matter is responsible for 54% of the variance

Site accounts for 88% of the variance

Season precip > Irrigation

Model Selection

Bioenv

Bioenv is an iterative procedure that links environmental variables to community structure by seeking the best subset of environmental variables that explains community structure

Info:

BIOENV

“The function calculates a community dissimilarity matrix using vegdist. Then it selects all possible subsets of environmental variables, scales the variables, and calculates Euclidean distances for this subset using dist.

The function finds the correlation between community dissimilarities and environmental distances, and for each size of subsets, saves the best result. There are $2^p - 1$ subsets of p variables, and an exhaustive search may take a very, very, very long time (parameter `upto` offers a partial relief)."

Mantel test

"Mantel statistic is simply a correlation between entries of two dissimilarity matrices (some use cross products, but these are linearly related). However, the significance cannot be directly assessed, because there are $N(N-1)/2$ entries for just N observations. Mantel developed asymptotic test, but here we use permutations of N rows and columns of dissimilarity matrix."

```
OTU_nifH_trim<- as(otu_table(physeq_nifH_ord ), "matrix")
```

```
#Transpose the data to have sample names on rows
abund_tablenifH<-t(OTU_nifH_trim)
```

```
nrow(abund_tablenifH)
```

```
## [1] 54
```

```
setdiff(rownames(meta3), rownames(abund_tablenifH))
```

```
## character(0)
```

Our meta data and sample data match with 0 difference in rownames

Will use parallel processing to speed up calculations

```
#First detect amount of cores available
detectCores()
```

```
## [1] 12
```

```
#get bray-curtis distance
```

```
abund_distnifH<-vegdist(abund_tablenifH, method = "bray")
```

```
#make bioenv model against whole meta table with ".", and use gower metric to measure distance in order
```

```
nifH.bioenv <- bioenv(abund_distnifH ~ ., meta3, index = "bray", method = "pearson",
                     metric = "gower", upto = 7, parallel = 10)
```

```
## Studying 1285623 of 134217727 possible subsets (this may take time...)
```

```
summary(nifH.bioenv)
```

```
##                                     size
## Site                             1
## Site Organic_Matter               2
## Site Organic_Matter Nitrate_Nitrite 3
## Site Organic_Matter Av_Phosphorus Zinc 4
## Site Organic_Matter Nitrate_Nitrite Av_Phosphorus Zinc 5
## Site Tillage Organic_Matter Av_Potassium Cobalt Zinc 6
## Site Tillage Organic_Matter Nitrate_Nitrite Ammonia Av_Potassium Zinc 7
##                                     correlation
## Site                             0.5968
## Site Organic_Matter               0.6068
## Site Organic_Matter Nitrate_Nitrite 0.6195
## Site Organic_Matter Av_Phosphorus Zinc 0.6128
## Site Organic_Matter Nitrate_Nitrite Av_Phosphorus Zinc 0.6062
## Site Tillage Organic_Matter Av_Potassium Cobalt Zinc 0.5995
## Site Tillage Organic_Matter Nitrate_Nitrite Ammonia Av_Potassium Zinc 0.5942
```

```
nifH.bioenvdist<-bioenvdist(nifH.bioenv, which = "best")
mantel(nifH.bioenvdist, abund_distnifH)
```

```
##
## Mantel statistic based on Pearson's product-moment correlation
##
## Call:
## mantel(xdis = nifH.bioenvdist, ydis = abund_distnifH)
##
## Mantel statistic r: 0.6195
##      Significance: 0.001
##
## Upper quantiles of permutations (null model):
##      90%    95%   97.5%    99%
## 0.0467 0.0616 0.0786 0.0879
## Permutation: free
## Number of permutations: 999
```

```
nifH.bioenv$whichbest
```

```
## [1] 3
```

The best model from the initial BioEnv model selection shows that the Site explains 59% of the variation in the bray-curitis distance. But the best model explains 63% of the variation when Organic Matter and nitritie and nitrate are added in.

Try BioENV without site, this will favor farm managment variables since they are nested within.

Remove site from meta table

```
meta4<-meta3[, -c(1)]
```

Rerun BioEnv

```
nifH.bioenv.site.na <- bioenv(abund_distnifH ~ ., meta4, index = "bray",  
                             method = "pearson", metric = "gower", upto = 7, parallel = 10)
```

```
## Studying 971711 of 67108863 possible subsets (this may take time...)
```

```
summary(nifH.bioenv.site.na)
```

```
##                                                                 size  
## Tillage                                                                 1  
## Tillage Organic_Matter                                                2  
## Tillage Organic_Matter Manganese                                       3  
## total_precip_irr Tillage Organic_Matter Zinc                          4  
## Tillage prev_crop Organic_Matter Av_Potassium Zinc                    5  
## Tillage prev_crop Organic_Matter Ammonia Av_Potassium Zinc            6  
## Tillage prev_crop Organic_Matter Ammonia Av_Potassium Manganese Zinc  7  
##                                                                 correlation  
## Tillage                                                                 0.3798  
## Tillage Organic_Matter                                                0.4868  
## Tillage Organic_Matter Manganese                                       0.5098  
## total_precip_irr Tillage Organic_Matter Zinc                          0.5263  
## Tillage prev_crop Organic_Matter Av_Potassium Zinc                    0.5328  
## Tillage prev_crop Organic_Matter Ammonia Av_Potassium Zinc            0.5363  
## Tillage prev_crop Organic_Matter Ammonia Av_Potassium Manganese Zinc  0.5395
```

```
nifH.bioenvdist.site.na<-bioenvdist(nifH.bioenv.site.na, which = "best")  
mantel(nifH.bioenvdist.site.na, abund_distnifH)
```

```
##  
## Mantel statistic based on Pearson's product-moment correlation  
##  
## Call:  
## mantel(xdis = nifH.bioenvdist.site.na, ydis = abund_distnifH)  
##  
## Mantel statistic r: 0.5395  
##      Significance: 0.001  
##  
## Upper quantiles of permutations (null model):  
##   90%   95%  97.5%   99%  
## 0.0446 0.0607 0.0762 0.0953  
## Permutation: free  
## Number of permutations: 999
```

```
nifH.bioenv.site.na$whichbest
```

```
## [1] 7
```

When Site is removed from the model selection we can see that there is still a bias for the Farm management variables because they are nested with the site factor. But we do see more come out to build a bigger model. With Tillage + prev_crop + Organic_Matter + Nitrate_Nitrite + Ammonia + Av_Potassium + Manganese + Zinc.

Bioenv with chem data only

```
nifH.bioenv.chem <- bioenv(abund_tablenifH ~ Organic_Matter + Moisture_Content +
  Nitrate_Nitrite + Ammonia + Av_Phosphorus + Av_Potassium +
  pH + Barium + Calcium + Cobalt + Copper + Iron + Magnesium +
  Manganese + Nickel + Phosphorus + Sulfur + Zinc, meta3,
  index = "bray", method = "pearson", metric = "gower",
  upto = 7, parallel = 10)
```

```
## Studying 63003 of 262143 possible subsets (this may take time...)
```

```
summary(nifH.bioenv.chem)
```

```
##
## size
## Organic_Matter 1
## Organic_Matter Zinc 2
## Organic_Matter Av_Phosphorus Zinc 3
## Organic_Matter Nitrate_Nitrite Av_Phosphorus Zinc 4
## Organic_Matter Nitrate_Nitrite Av_Phosphorus Manganese Zinc 5
## Organic_Matter Nitrate_Nitrite Av_Phosphorus Av_Potassium Barium Zinc 6
## Organic_Matter Nitrate_Nitrite Ammonia Av_Phosphorus Av_Potassium Barium Zinc 7
## correlation
## Organic_Matter 0.3716
## Organic_Matter Zinc 0.4439
## Organic_Matter Av_Phosphorus Zinc 0.4922
## Organic_Matter Nitrate_Nitrite Av_Phosphorus Zinc 0.4894
## Organic_Matter Nitrate_Nitrite Av_Phosphorus Manganese Zinc 0.4819
## Organic_Matter Nitrate_Nitrite Av_Phosphorus Av_Potassium Barium Zinc 0.4872
## Organic_Matter Nitrate_Nitrite Ammonia Av_Phosphorus Av_Potassium Barium Zinc 0.4848
```

```
nifH.bioenvdist.chem<-bioenvdist(nifH.bioenv.chem, which = "best")
mantel(nifH.bioenvdist.chem, abund_distnifH)
```

```
##
## Mantel statistic based on Pearson's product-moment correlation
##
## Call:
## mantel(xdis = nifH.bioenvdist.chem, ydis = abund_distnifH)
##
## Mantel statistic r: 0.4922
## Significance: 0.001
##
## Upper quantiles of permutations (null model):
## 90% 95% 97.5% 99%
## 0.0506 0.0662 0.0826 0.0932
## Permutation: free
## Number of permutations: 999
```

```
nifH.bioenv.chem$whichbest
```

```
## [1] 3
```

The fourth model Organic Matter, Nitrate, Phosphorus and Zinc is the best explaining 56% of the variance. Which is the same as before without the location variables.

CCA/ordistep model selection

CCA model selection uses a procedure to take the a constrained distance ordination with the complete model and compare it with a unconstrained model. The model starts with no variables then adds variables that make the best model. These models can then be plotted against the same ordination space as vectors.

Ordistep use Akaike information criterion (AIC) which is a estimator of relative quality of the models. AIC is relative to models you are comparing when you compare two models the one with the lower AIC is favored.

```
m1_nifH <- cca(abund_tablenifH ~ ., meta3)
m0_nifH <- cca(abund_tablenifH ~ 1, meta3)
m1_nifH

## Call: cca(formula = abund_tablenifH ~ Site + Pea_variety + Plot +
## season_precip + irrigation + total_precip_irr + Tillage + prev_crop
## + grain_yield + Organic_Matter + Moisture_Content +
## Nitrate_Nitrite + Ammonia + Av_Phosphorus + Av_Potassium + pH +
## Barium + Calcium + Cobalt + Copper + Iron + Magnesium + Manganese
## + Nickel + Phosphorus + Sulfur + Zinc, data = meta3)
##
##              Inertia Proportion Rank
## Total              9.074         1.000
## Constrained        6.796         0.749   33
## Unconstrained      2.277         0.251   20
## Inertia is scaled Chi-square
## Some constraints were aliased because they were collinear (redundant)
##
## Eigenvalues for constrained axes:
##   CCA1   CCA2   CCA3   CCA4   CCA5   CCA6   CCA7   CCA8   CCA9   CCA10
## 0.8027 0.6026 0.5827 0.5207 0.4910 0.4632 0.3482 0.3227 0.2719 0.2619
##   CCA11  CCA12  CCA13  CCA14  CCA15  CCA16  CCA17  CCA18  CCA19  CCA20
## 0.2525 0.1977 0.1824 0.1742 0.1564 0.1370 0.1242 0.1143 0.1030 0.0979
##   CCA21  CCA22  CCA23  CCA24  CCA25  CCA26  CCA27  CCA28  CCA29  CCA30
## 0.0751 0.0699 0.0663 0.0567 0.0515 0.0488 0.0444 0.0370 0.0353 0.0326
##   CCA31  CCA32  CCA33
## 0.0283 0.0241 0.0193
##
## Eigenvalues for unconstrained axes:
##    CA1    CA2    CA3    CA4    CA5    CA6    CA7    CA8
## 0.27474 0.23444 0.19971 0.16799 0.14878 0.14661 0.14098 0.12503
## (Showing 8 of 20 unconstrained eigenvalues)
```

```
m0_nifH
```



```
## Call: cca(formula = abund_tablenifH ~ 1, data = meta3)
##
##              Inertia Rank
## Total          9.074
## Unconstrained  9.074   53
## Inertia is scaled Chi-square
##
## Eigenvalues for unconstrained axes:
##   CA1   CA2   CA3   CA4   CA5   CA6   CA7   CA8
## 0.8150 0.6423 0.6049 0.5542 0.5252 0.4870 0.4435 0.3923
## (Showing 8 of 53 unconstrained eigenvalues)
```

Ordistep

```
model_nifH <-ordistep(m0_nifH, scope=formula(m1_nifH))
```

```
model_nifH$anova
```

	Df	AIC	F	Pr(>F)
+ Site	8	856.3138	3.617355	0.005
+ Nitrate_Nitrite	1	855.3388	2.492073	0.010

Everytime a variable is added the AIC is lowered favoring the new model. This goes till it stops.

cca without site

```
m1_nifH_site_na <- cca(abund_tablenifH ~ ., meta4)
m0_nifH_site_na <- cca(abund_tablenifH ~ 1, meta4)
m1_nifH_site_na
```

```
## Call: cca(formula = abund_tablenifH ~ Pea_variety + Plot +
## season_precip + irrigation + total_precip_irr + Tillage + prev_crop
## + grain_yield + Organic_Matter + Moisture_Content +
## Nitrate_Nitrite + Ammonia + Av_Phosphorus + Av_Potassium + pH +
## Barium + Calcium + Cobalt + Copper + Iron + Magnesium + Manganese
## + Nickel + Phosphorus + Sulfur + Zinc, data = meta4)
##
##              Inertia Proportion Rank
## Total          9.0738      1.0000
## Constrained    6.6182      0.7294   32
## Unconstrained  2.4557      0.2706   21
## Inertia is scaled Chi-square
## Some constraints were aliased because they were collinear (redundant)
##
## Eigenvalues for constrained axes:
##   CCA1   CCA2   CCA3   CCA4   CCA5   CCA6   CCA7   CCA8   CCA9   CCA10
## 0.8024 0.6003 0.5814 0.5066 0.4808 0.4506 0.3469 0.3226 0.2675 0.2609
##   CCA11  CCA12  CCA13  CCA14  CCA15  CCA16  CCA17  CCA18  CCA19  CCA20
```

```
## 0.2103 0.1972 0.1822 0.1609 0.1540 0.1356 0.1220 0.1142 0.1019 0.0810
## CCA21 CCA22 CCA23 CCA24 CCA25 CCA26 CCA27 CCA28 CCA29 CCA30
## 0.0744 0.0663 0.0633 0.0522 0.0512 0.0468 0.0420 0.0355 0.0333 0.0283
## CCA31 CCA32
## 0.0263 0.0193
##
## Eigenvalues for unconstrained axes:
## CA1 CA2 CA3 CA4 CA5 CA6 CA7 CA8
## 0.28403 0.24996 0.21164 0.18851 0.16085 0.14877 0.14524 0.12522
## (Showing 8 of 21 unconstrained eigenvalues)
```

```
m0_nifH_site_na
```

```
## Call: cca(formula = abund_tablenifH ~ 1, data = meta4)
##
## Inertia Rank
## Total 9.074
## Unconstrained 9.074 53
## Inertia is scaled Chi-square
##
## Eigenvalues for unconstrained axes:
## CA1 CA2 CA3 CA4 CA5 CA6 CA7 CA8
## 0.8150 0.6423 0.6049 0.5542 0.5252 0.4870 0.4435 0.3923
## (Showing 8 of 53 unconstrained eigenvalues)
```

Ordistep

```
model_nifH_site_na <- ordistep(m0_nifH_site_na, scope=formula(m1_nifH_site_na))
```

```
model_nifH_site_na$anova
```

	Df	AIC	F	Pr(>F)
+ Tillage	2	864.3021	3.436408	0.005
+ prev_crop	3	860.8917	3.045976	0.005
+ irrigation	1	858.7091	3.785031	0.005
+ Ammonia	1	858.2541	2.139600	0.005
+ season_precip	1	857.5351	2.323828	0.005
+ Nitrate_Nitrite	1	856.6017	2.456277	0.005
+ grain_yield	1	856.1815	1.971028	0.010

```
m1_nifH_cca_chem <- cca(abund_tablenifH ~ Organic_Matter + Moisture_Content + Nitrate_Nitrite + Ammonia
m0_nifH_cca_chem <- cca(abund_tablenifH ~ 1, meta3)
m1_nifH_cca_chem
```

```
## Call: cca(formula = abund_tablenifH ~ Organic_Matter +
## Moisture_Content + Nitrate_Nitrite + Ammonia + Av_Phosphorus +
## Av_Potassium + pH + Barium + Calcium + Cobalt + Copper + Iron +
## Magnesium + Manganese + Nickel + Phosphorus + Sulfur + Zinc, data
## = meta3)
```

```
##
##              Inertia Proportion Rank
## Total          9.0738      1.0000
## Constrained    4.6226      0.5094   18
## Unconstrained  4.4513      0.4906   35
## Inertia is scaled Chi-square
##
## Eigenvalues for constrained axes:
##   CCA1  CCA2  CCA3  CCA4  CCA5  CCA6  CCA7  CCA8  CCA9  CCA10
## 0.7877 0.5566 0.5201 0.4856 0.4519 0.3928 0.3003 0.1899 0.1705 0.1467
##   CCA11 CCA12 CCA13 CCA14 CCA15 CCA16 CCA17 CCA18
## 0.1313 0.1064 0.0846 0.0741 0.0684 0.0621 0.0500 0.0436
##
## Eigenvalues for unconstrained axes:
##   CA1  CA2  CA3  CA4  CA5  CA6  CA7  CA8
## 0.4395 0.3474 0.3067 0.2537 0.2382 0.2336 0.2226 0.2018
## (Showing 8 of 35 unconstrained eigenvalues)
```

```
m0_nifH_cca_chem
```

```
## Call: cca(formula = abund_tablenifH ~ 1, data = meta3)
##
##              Inertia Rank
## Total          9.074
## Unconstrained  9.074   53
## Inertia is scaled Chi-square
##
## Eigenvalues for unconstrained axes:
##   CA1  CA2  CA3  CA4  CA5  CA6  CA7  CA8
## 0.8150 0.6423 0.6049 0.5542 0.5252 0.4870 0.4435 0.3923
## (Showing 8 of 53 unconstrained eigenvalues)
```

```
model_nifH_cca_chem <- ordistep(m0_nifH_cca_chem, scope=formula(m1_nifH_cca_chem))
```

```
model_nifH_cca_chem$anova
```

	Df	AIC	F	Pr(>F)
+ Organic_Matter	1	864.9180	4.217224	0.005
+ Nitrate_Nitrite	1	863.6253	3.206498	0.005
+ Av_Potassium	1	862.6412	2.840851	0.005
+ Av_Phosphorus	1	861.7903	2.656463	0.005
+ Manganese	1	861.1619	2.394125	0.005
+ Barium	1	860.3621	2.501186	0.005
+ Sulfur	1	859.9429	2.107586	0.005
+ Magnesium	1	858.8295	2.670746	0.005
+ Phosphorus	1	858.4552	1.977791	0.005

RDA is a linear cca

```
m1_nifH_rda <- rda(abund_tablenifH ~ ., meta3)
m0_nifH_rda <- rda(abund_tablenifH ~ 1, meta3)
m1_nifH_rda
```

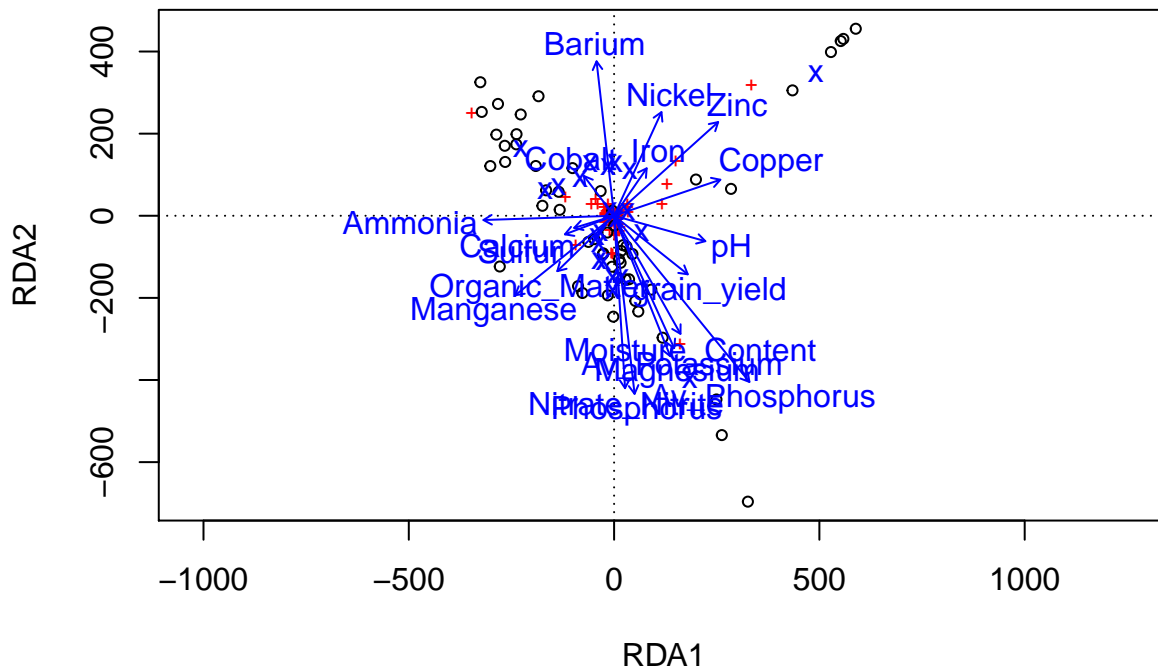
```
## Call: rda(formula = abund_tablenifH ~ Site + Pea_variety + Plot +
## season_precip + irrigation + total_precip_irr + Tillage + prev_crop
## + grain_yield + Organic_Matter + Moisture_Content +
## Nitrate_Nitrite + Ammonia + Av_Phosphorus + Av_Potassium + pH +
## Barium + Calcium + Cobalt + Copper + Iron + Magnesium + Manganese
## + Nickel + Phosphorus + Sulfur + Zinc, data = meta3)
##
##              Inertia Proportion Rank
## Total          1.331e+11  1.000e+00
## Constrained    1.046e+11  7.862e-01   33
## Unconstrained  2.846e+10  2.138e-01   20
## Inertia is variance
## Some constraints were aliased because they were collinear (redundant)
##
## Eigenvalues for constrained axes:
##      RDA1      RDA2      RDA3      RDA4      RDA5      RDA6
## 17548667644 16499650004 12977757804 11589506943 7631119218 6761513528
##      RDA7      RDA8      RDA9      RDA10     RDA11     RDA12
## 5550551834 5004470215 4122249247 3318880699 2680128188 1968062864
##      RDA13     RDA14     RDA15     RDA16     RDA17     RDA18
## 1497032684 1192899104 1042107947 861647889 698124705 588564180
##      RDA19     RDA20     RDA21     RDA22     RDA23     RDA24
## 508094041 418009285 365033164 324702400 252472753 207429340
##      RDA25     RDA26     RDA27     RDA28     RDA29     RDA30
## 184825179 169758166 150199224 115815986 107173357 96024376
##      RDA31     RDA32     RDA33
## 84736502 60931011 58747462
##
## Eigenvalues for unconstrained axes:
##      PC1      PC2      PC3      PC4      PC5      PC6
## 5871852873 4402692918 4002911334 3724051954 2047287739 1519664997
##      PC7      PC8
## 1228123308 998854527
## (Showing 8 of 20 unconstrained eigenvalues)
```

```
m0_nifH_rda
```

```
## Call: rda(formula = abund_tablenifH ~ 1, data = meta3)
##
##              Inertia Rank
## Total          1.331e+11
## Unconstrained  1.331e+11   53
## Inertia is variance
##
## Eigenvalues for unconstrained axes:
##      PC1      PC2      PC3      PC4      PC5      PC6
## 19425920664 18485720261 15308306751 13963841933 10802018954 8105770060
##      PC7      PC8
## 7008929496 6473486029
```

```
## (Showing 8 of 53 unconstrained eigenvalues)
```

```
plot(m1_nifH_rda)
```



```
model_rda_nifH <-ordiR2step(m0_nifH_rda, scope=formula(m1_nifH_rda))
```

```
model_rda_nifH$anova
```

	R2.adj	Df	AIC	F	Pr(>F)
+ Site	0.4096331	8	1362.869	5.596835	0.002
	0.4334037	NA	NA	NA	NA

Only site appears in the complete RDA model. This is a liner based method and might not work well with out data.

Chemistry only RDA model

```
m1_nifH_rda_chem <- rda(abund_tablenifH ~ Organic_Matter + Moisture_Content + Nitrate_Nitrite + Ammonia
m0_nifH_rda_chem <- rda(abund_tablenifH ~ 1, meta3)
m1_nifH_rda_chem
```

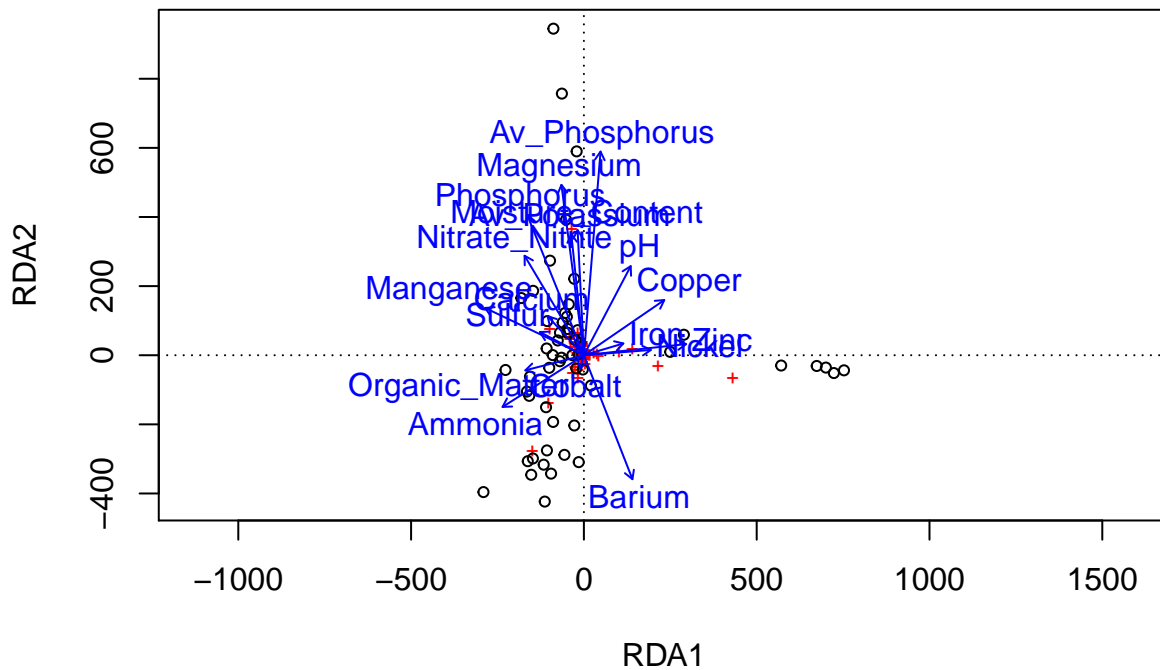
```
## Call: rda(formula = abund_tablenifH ~ Organic_Matter +
## Moisture_Content + Nitrate_Nitrite + Ammonia + Av_Phosphorus +
## Av_Potassium + pH + Barium + Calcium + Cobalt + Copper + Iron +
## Magnesium + Manganese + Nickel + Phosphorus + Sulfur + Zinc, data
## = meta3)
##
```

```
##              Inertia Proportion Rank
## Total          1.331e+11  1.000e+00
## Constrained    7.463e+10  5.607e-01  18
## Unconstrained  5.846e+10  4.393e-01  35
## Inertia is variance
##
## Eigenvalues for constrained axes:
##      RDA1      RDA2      RDA3      RDA4      RDA5      RDA6
## 16122227703 13523241448 11094526423 9005789057 5921938961 5359338168
##      RDA7      RDA8      RDA9      RDA10     RDA11     RDA12
##  3780506525 3080678651 1894819093 1508918418  940183925  624720502
##      RDA13     RDA14     RDA15     RDA16     RDA17     RDA18
##   471289405  382996377  289104096  239329825  227754724  162233619
##
## Eigenvalues for unconstrained axes:
##      PC1      PC2      PC3      PC4      PC5      PC6
## 11949328777  7624704552  6257652081  5543387518  4129403632  3255191916
##      PC7      PC8
##  3002483572  2695244220
## (Showing 8 of 35 unconstrained eigenvalues)
```

```
m0_nifH_rda_chem
```

```
## Call: rda(formula = abund_tablenifH ~ 1, data = meta3)
##
##              Inertia Rank
## Total          1.331e+11
## Unconstrained  1.331e+11  53
## Inertia is variance
##
## Eigenvalues for unconstrained axes:
##      PC1      PC2      PC3      PC4      PC5      PC6
## 19425920664 18485720261 15308306751 13963841933 10802018954 8105770060
##      PC7      PC8
##  7008929496  6473486029
## (Showing 8 of 53 unconstrained eigenvalues)
```

```
plot(m1_nifH_rda_chem)
```



```
model_rda_chem_nifH <- ordiR2step(m0_nifH_rda_chem, scope=formula(m1_nifH_rda_chem))
```

```
model_rda_chem_nifH$anova
```

	R2.adj	Df	AIC	F	Pr(>F)
+ Av_Phosphorus	0.0580227	1	1381.907	4.264624	0.002
+ Av_Potassium	0.1012829	1	1380.320	3.503049	0.002
+ Organic_Matter	0.1424550	1	1378.719	3.448590	0.002
+ Barium	0.1852360	1	1376.864	3.625360	0.002
+ Manganese	0.2506562	1	1373.231	5.277867	0.002
+ Iron	0.2865280	1	1371.445	3.413331	0.002
+ Sulfur	0.3191470	1	1369.757	3.251722	0.002
+ Magnesium	0.3325579	1	1369.496	1.924280	0.030
	0.3348198	NA	NA	NA	NA

Our RDA chem model can explain 30% of the variance in the bacterial community.

CAP Ordination model building

canonical analysis of principal coordinates (CAP) is similar to RDA but allows for non-euclidian dissimilarity like Bray-Curtis which we have been using.

<https://esajournals.onlinelibrary.wiley.com/doi/epdf/10.1890/0012-9658%282003%29084%5B0511%3ACAOPCA%5D2.0.CO%3B2>

Guide for the CAP Ordination

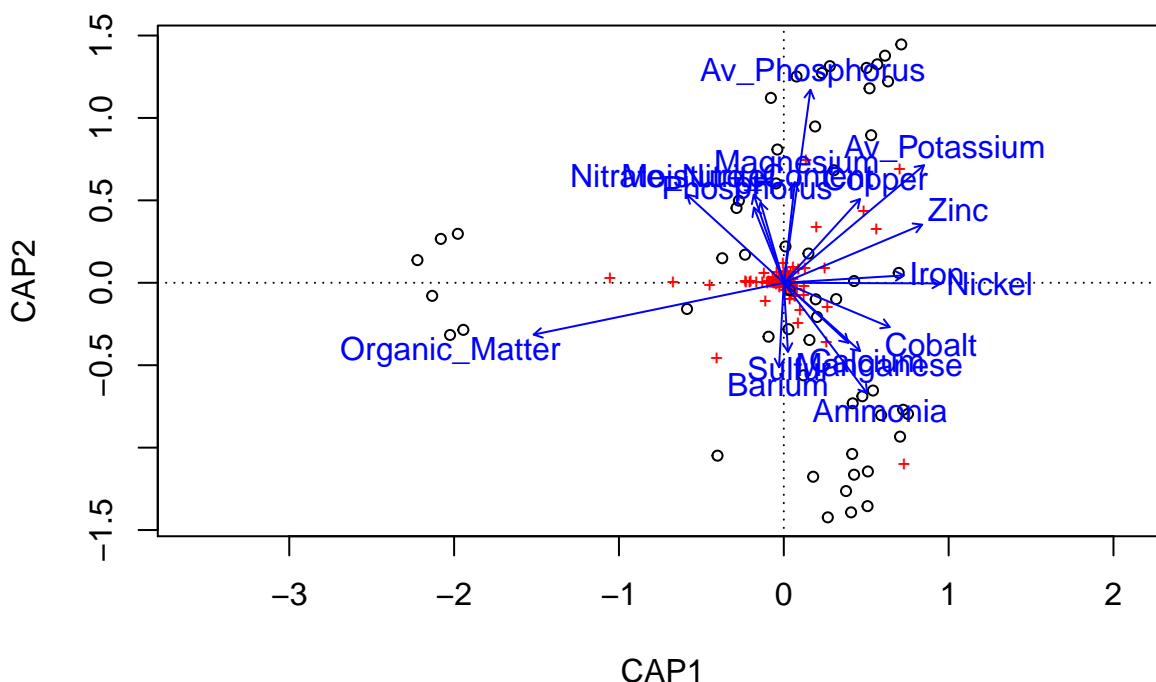
<https://quantpalaeo.wordpress.com/2014/04/14/variance-inflation-factors-and-ordination-model-selection/>

If the VIF of a predictor is high, it indicates that that predictor is highly correlated with other predictors, it contains little or no unique information, and there is redundancy in the set of predictors.

STEPS

- 1)Generate a constrained ordination with all available predictors.
- 2)Calculate the VIF of each variable.
- 3)If any variable has a VIF over a threshold (typically 10), drop the variable with the highest VIF
- 4)Repeat until all remaining variables have a VIF below the threshold.

```
m1_nifH_cap_chem<- capscale(abund_tablenifH ~ Organic_Matter + Moisture_Content + Nitrate_Nitrite + Ammonia
m0_nifH_cap_chem<- capscale(abund_distnifH ~ 1, meta3)
plot(m1_nifH_cap_chem)
```



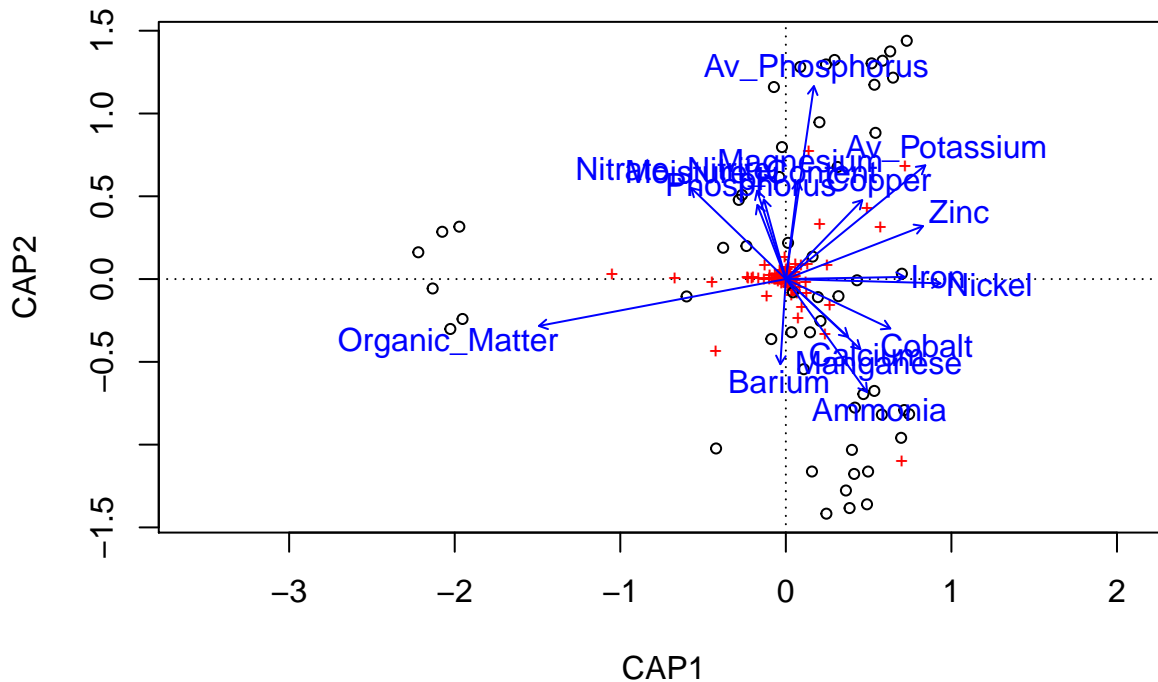
```
vif.cca(m1_nifH_cap_chem)
```

##	Organic_Matter	Moisture_Content	Nitrate_Nitrite	Ammonia
##	15.457094	8.111272	4.360780	4.369015
##	Av_Phosphorus	Av_Potassium	pH	Barium
##	11.821766	13.816708	19.007192	34.110255
##	Calcium	Cobalt	Copper	Iron
##	69.959837	20.726173	37.675483	55.160314
##	Magnesium	Manganese	Nickel	Phosphorus


```
##      38.155344      25.416138      39.209635      37.072036
##      Sulfur      Zinc
##      126.793579      83.408013
```

Removing Sulfur

```
m1_nifH_cap_chem_1<- capscale(abund_tablenifH ~ Organic_Matter + Moisture_Content + Nitrate_Nitrite + Ammonia)
m0_nifH_cap_chem<- capscale(abund_distnifH ~ 1, meta3)
plot(m1_nifH_cap_chem_1)
```

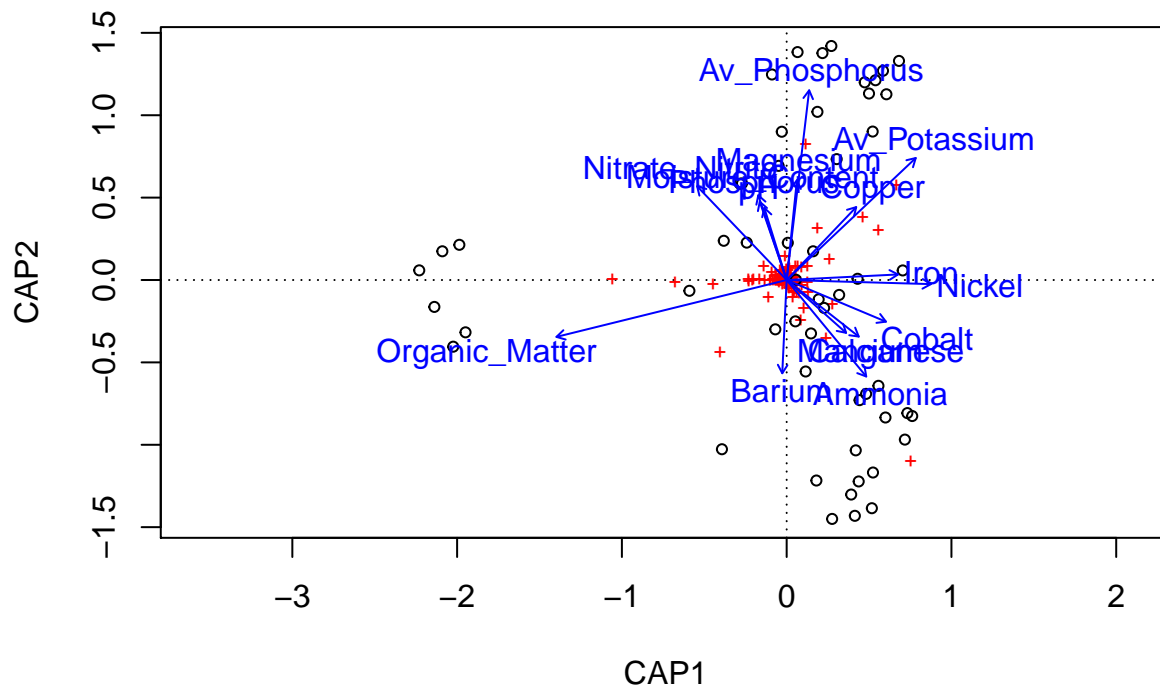


```
vif.cca(m1_nifH_cap_chem_1)
```

```
##      Organic_Matter Moisture_Content Nitrate_Nitrite      Ammonia
##      14.498699      6.988162      3.713414      4.309993
##      Av_Phosphorus      Av_Potassium      pH      Barium
##      11.350534      12.916393      16.155227      31.533526
##      Calcium      Cobalt      Copper      Iron
##      9.761153      15.971211      35.302426      53.661315
##      Magnesium      Manganese      Nickel      Phosphorus
##      31.551213      17.840002      36.191208      28.682249
##      Zinc
##      80.556370
```

Removing Zinc

```
m1_nifH_cap_chem_2<- capscale(abund_tablenifH ~ Organic_Matter + Moisture_Content + Nitrate_Nitrite + Ammonia)
m0_nifH_cap_chem<- capscale(abund_distnifH ~ 1, meta3)
plot(m1_nifH_cap_chem_2)
```

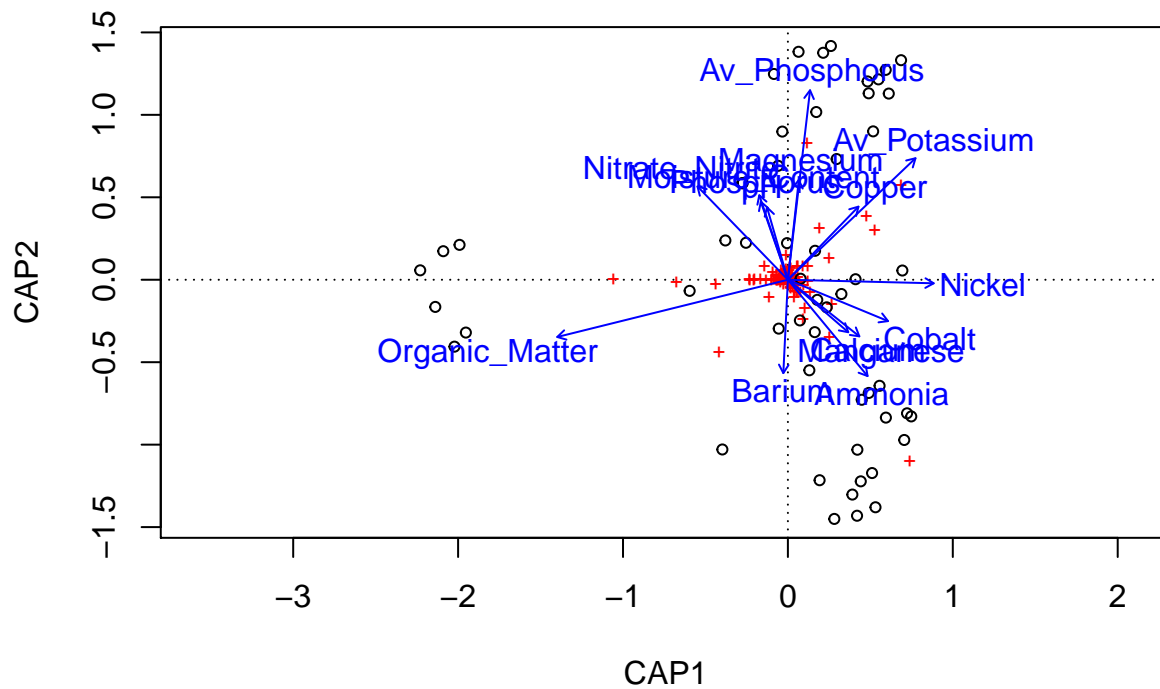


```
vif.cca(m1_nifH_cap_chem_2)
```

##	Organic_Matter	Moisture_Content	Nitrate_Nitrite	Ammonia
##	12.646738	5.847957	3.672388	4.309848
##	Av_Phosphorus	Av_Potassium	pH	Barium
##	11.225914	10.884874	15.913142	18.081149
##	Calcium	Cobalt	Copper	Iron
##	9.388663	14.920297	31.854122	41.065111
##	Magnesium	Manganese	Nickel	Phosphorus
##	31.536509	16.206354	29.523059	28.553042

Removing Iron

```
m1_nifH_cap_chem_3<- capscale(abund_tablenifH ~ Organic_Matter + Moisture_Content + Nitrate_Nitrite + Ammonia, meta3)
m0_nifH_cap_chem<- capscale(abund_distnifH ~ 1, meta3)
plot(m1_nifH_cap_chem_3)
```

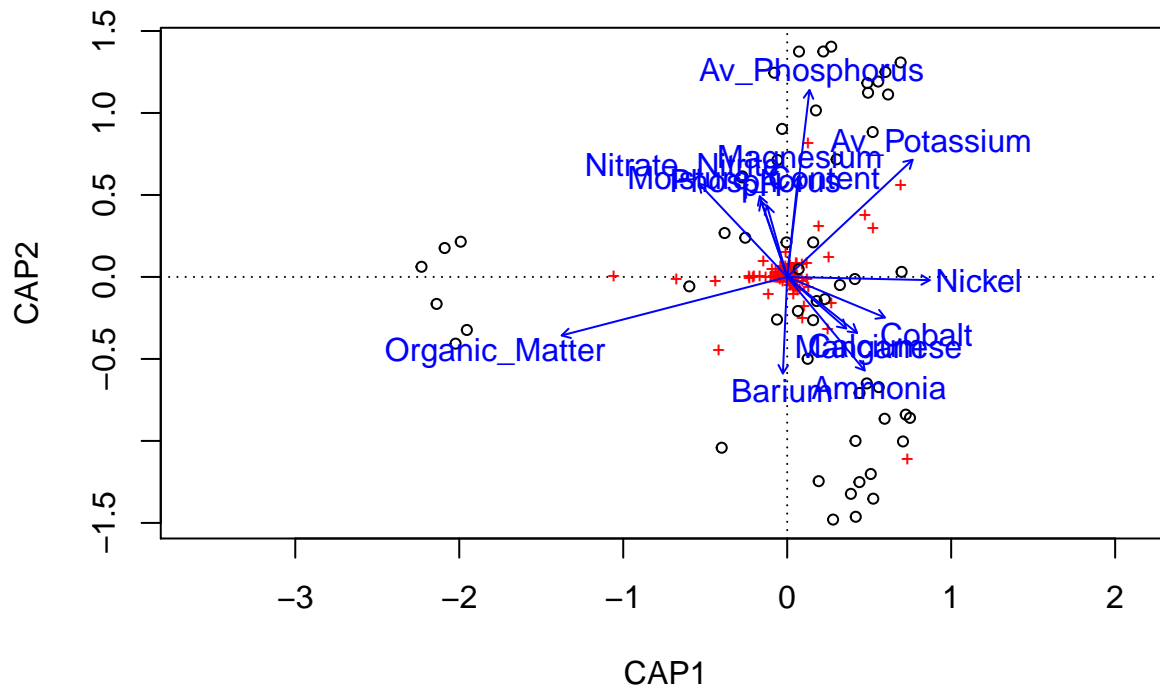


```
vif.cca(m1_nifH_cap_chem_3)
```

##	Organic_Matter	Moisture_Content	Nitrate_Nitrite	Ammonia
##	10.243305	5.564319	3.533075	4.304512
##	Av_Phosphorus	Av_Potassium	pH	Barium
##	10.829226	10.743500	14.653678	17.693848
##	Calcium	Cobalt	Copper	Magnesium
##	6.990169	13.202442	30.558543	29.869859
##	Manganese	Nickel	Phosphorus	
##	14.958783	24.569283	17.904712	

Removing Copper

```
m1_nifH_cap_chem_4<- capscale(abund_tablenifH ~ Organic_Matter + Moisture_Content + Nitrate_Nitrite + Ammonia, meta3)
m0_nifH_cap_chem<- capscale(abund_distnifH ~ 1, meta3)
plot(m1_nifH_cap_chem_4)
```

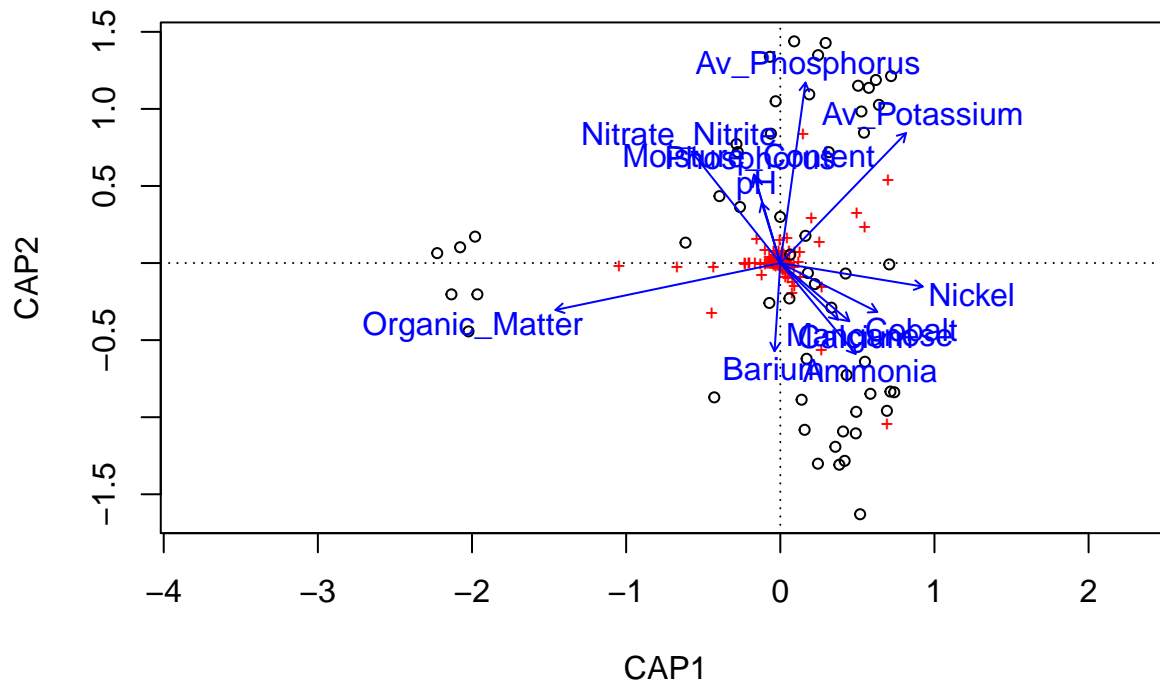


```
vif.cca(m1_nifH_cap_chem_4)
```

```
##      Organic_Matter Moisture_Content Nitrate_Nitrite      Ammonia
##      9.382412      5.403654      3.279764      4.290460
##      Av_Phosphorus      Av_Potassium      pH      Barium
##      9.817094      10.719638      14.623835      9.620284
##      Calcium      Cobalt      Magnesium      Manganese
##      6.023507      11.675665      22.563802      10.794157
##      Nickel      Phosphorus
##      21.922046      16.432692
```

Removing Magnesium

```
m1_nifH_cap_chem_5<- capscale(abund_tablenifH ~ Organic_Matter + Moisture_Content + Nitrate_Nitrite + Ammonia,
m0_nifH_cap_chem<- capscale(abund_distnifH ~ 1, meta3)
plot(m1_nifH_cap_chem_5)
```

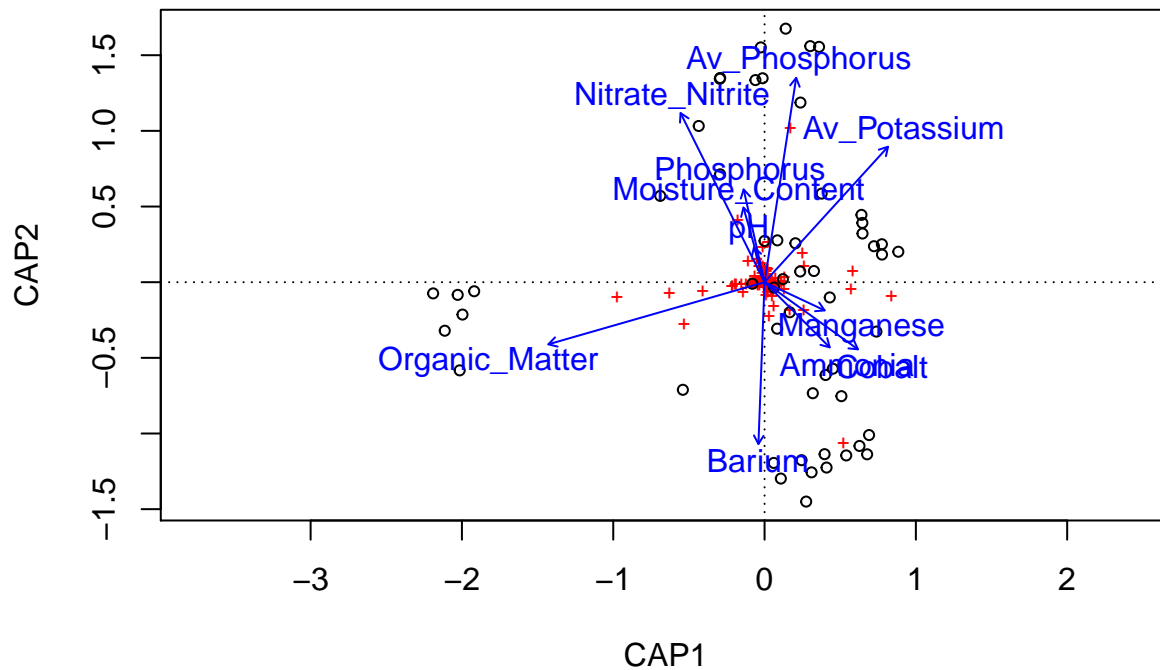


```
vif.cca(m1_nifH_cap_chem_5)
```

```
## Organic_Matter Moisture_Content Nitrate_Nitrite Ammonia
## 9.280442 5.335735 3.115965 4.278840
## Av_Phosphorus Av_Potassium pH Barium
## 9.280456 9.657791 12.581736 5.989495
## Calcium Cobalt Manganese Nickel
## 4.196095 11.661999 7.354337 18.320162
## Phosphorus
## 10.931904
```

Removing Nickel

```
m1_nifH_cap_chem_6<- capscale(abund_tablenifH ~ Organic_Matter + Moisture_Content + Nitrate_Nitrite + Ammonia)
m0_nifH_cap_chem<- capscale(abund_distnifH ~ 1, meta3)
plot(m1_nifH_cap_chem_6)
```

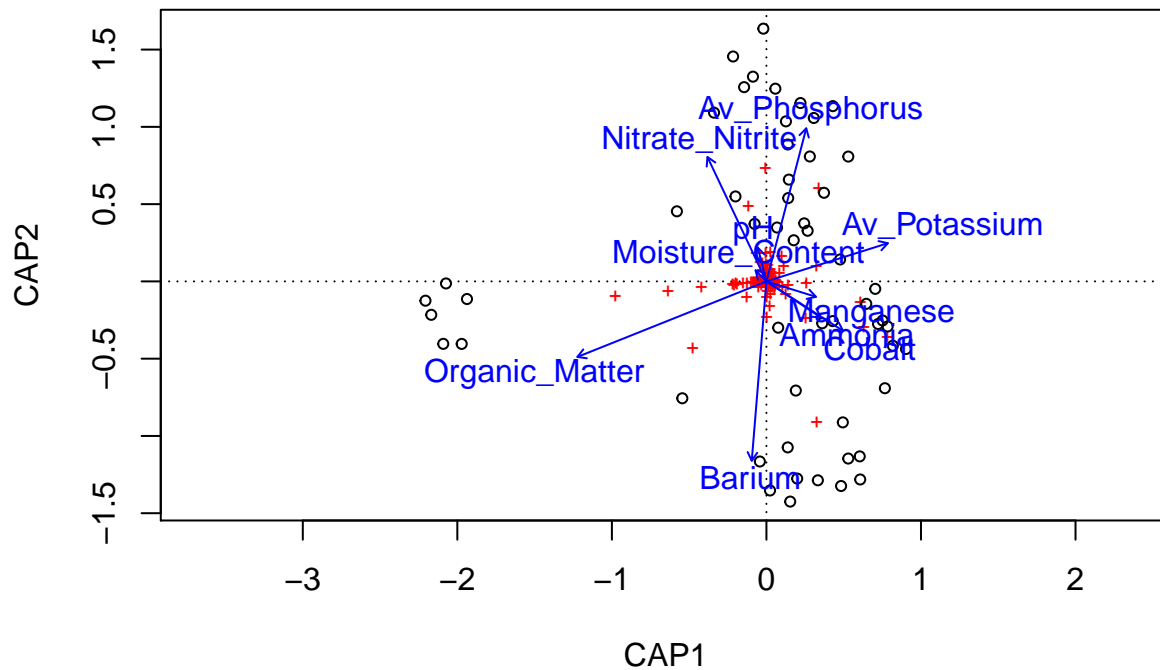


```
vif.cca(m1_nifH_cap_chem_6)
```

##	Organic_Matter	Moisture_Content	Nitrate_Nitrite	Ammonia
##	4.217221	5.071432	3.036605	4.024085
##	Av_Phosphorus	Av_Potassium	pH	Barium
##	6.845426	6.772406	9.848316	3.102538
##	Cobalt	Manganese	Phosphorus	
##	6.612077	5.670644	10.227267	

Removing Phosphorus as final covariant

```
m1_nifH_cap_chem_7<- capscale(abund_tablenifH ~ Organic_Matter + Moisture_Content + Nitrate_Nitrite + Ammonia)
m0_nifH_cap_chem<- capscale(abund_distnifH ~ 1, meta3)
plot(m1_nifH_cap_chem_7)
```



```
vif.cca(m1_nifH_cap_chem_7)
```

```
## Organic_Matter Moisture_Content Nitrate_Nitrite Ammonia
## 3.289053 4.848379 2.385440 3.637139
## Av_Phosphorus Av_Potassium pH Barium
## 6.842583 5.484280 7.125107 3.102513
## Cobalt Manganese
## 6.161856 5.169878
```

Every variable is under 10 will proceed to model building

```
model_cap_chem_nifH <-ordiR2step(m0_nifH_cap_chem, scope=formula(m1_nifH_cap_chem_7))
```

```
aov_model_cap_chem_nifH<-model_cap_chem_nifH$anova
aov_model_cap_chem_nifH
```

	R2.adj	Df	AIC	F	Pr(>F)
+ Organic_Matter	0.0902795	1	158.5307	6.053807	0.002
+ Barium	0.1477390	1	156.1156	4.345045	0.002
+ Av_Potassium	0.2066181	1	153.3646	4.598390	0.002
+ Manganese	0.2657762	1	150.3047	4.813408	0.002
+ Av_Phosphorus	0.3155079	1	147.6230	4.347257	0.002
+ Cobalt	0.3584885	1	145.2098	4.002428	0.002
+ Nitrate_Nitrite	0.3794547	1	144.4078	2.449920	0.006
	0.3935511	NA	NA	NA	NA

make table

```
capture.output(aov_model_cap_chem_nifH,file="aov_model_cap_chem_nifH.txt")
```

Constrained Ordination

["http://denefflab.github.io/MicrobeMiseq/demos/mothur_2_phyloseq.html#constrained_ordinations"](http://denefflab.github.io/MicrobeMiseq/demos/mothur_2_phyloseq.html#constrained_ordinations)

CCA with selected model from Chemical CCA model

```
# CCA ordinate
cca_ord_nifH <- ordinate(
  physeq = physeq_nifH_ord,
  method = "CCA",
  distance = abund_distnifH,
  formula = ~ Organic_Matter + Nitrate_Nitrite + Av_Potassium +
    Av_Phosphorus + Manganese + Barium + Sulfur + Magnesium )

# CCA plot
cca_plot_nifH <- plot_ordination(
  physeq = physeq_nifH_ord,
  ordination = cca_ord_nifH,
  color = "Site",
  axes = c(1,2)) +
  geom_point(aes(colour = Site), size = 3) +
  scale_color_manual(values = farm_col_paired)

# Now add the environmental variables as arrows
arrowmat_nifH_cca <- vegan::scores(cca_ord_nifH, display = "bp")

#Get appropriate scaling multiplier
mul<-vegan::ordiArrowMul(arrowmat_nifH_cca)

#Multiply biplot by scaling multiplier
arrowmat_nifH_cca_scale<-arrowmat_nifH_cca*3

# Add labels, make a data.frame
arrowdf_nifH_cca <- data.frame(labels = rownames(arrowmat_nifH_cca_scale), arrowmat_nifH_cca_scale)

# Define the arrow aesthetic mapping
arrow_map <- aes(xend = CCA1,
  yend = CCA2,
  x = 0,
  y = 0,
  shape = NULL,
  color = NULL,
  label = labels)

label_map <- aes(x = 1.3* CCA1,
  y = 1.3 * CCA2,
  shape = NULL,
  color = NULL,
```



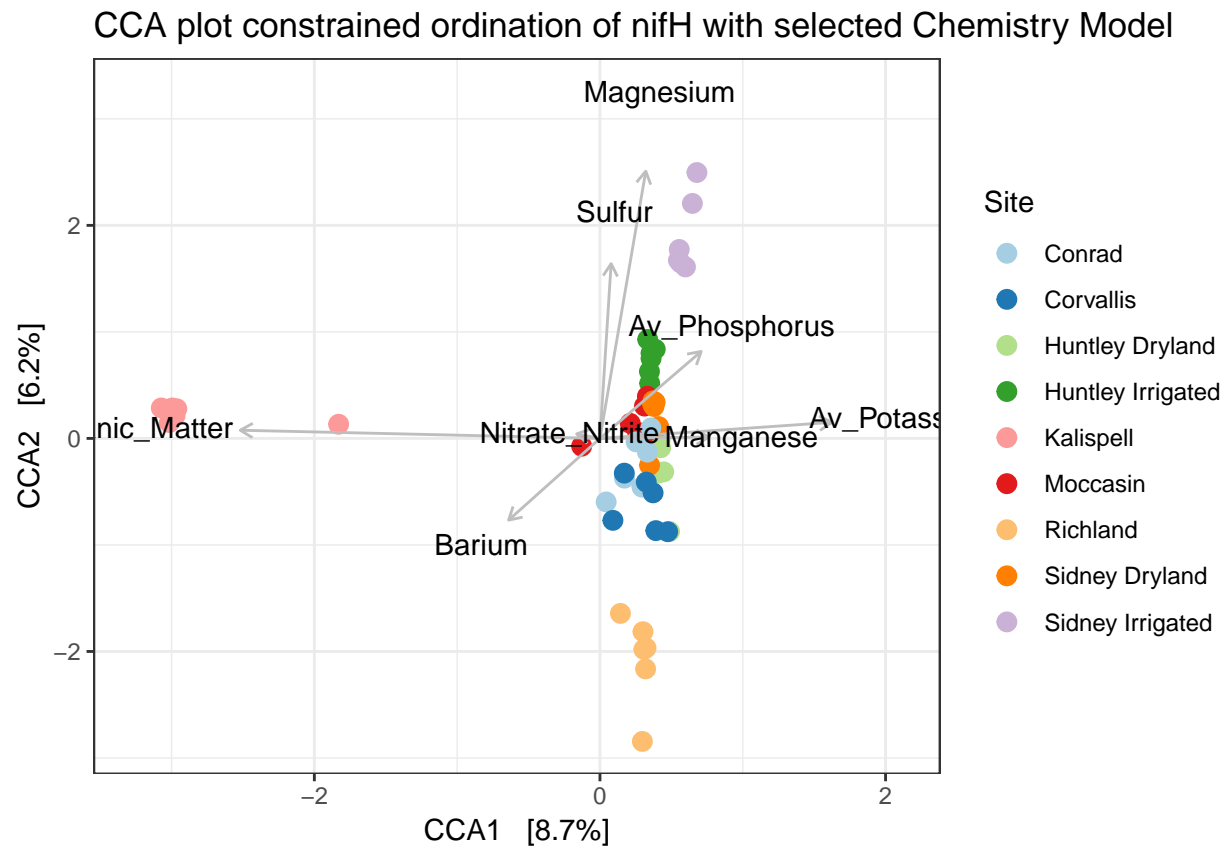
```

    label = labels)

arrowhead = arrow(length = unit(0.02, "npc"))

# Make a new graphic
cca_plot_nifH +
  geom_segment(
    mapping = arrow_map,
    size = .5,
    data = arrowdf_nifH_cca,
    color = "gray",
    arrow = arrowhead
  ) +
  geom_text(
    mapping = label_map,
    size = 4,
    data = arrowdf_nifH_cca,
    show.legend = FALSE
  ) +
  ggtitle("CCA plot constrained ordination of nifH with selected Chemistry Model") +
  theme_bw()

```



RDA with selected model from Chemical RDA model

```

# RDA ordinate
rda_ord_nifH <- ordinate(
  physeq = physeq_nifH_ord,
  method = "RDA",
  distance = abund_distnifH,
  formula = ~ Av_Phosphorus + Sulfur + Magnesium + Zinc + Av_Potassium + Manganese )

# RDA plot
rda_plot_nifH <- plot_ordination(
  physeq = physeq_nifH_ord,
  ordination = rda_ord_nifH,
  color = "Site",
  axes = c(1,2)) +
  geom_point(aes(colour = Site), size = 3) +
  scale_color_manual(values = farm_col_paired)

# Now add the environmental variables as arrows
arrowmat_nifH_rda <- vegan::scores(rda_ord_nifH, display = "bp")

#Get appropriate scaling multiplier
mul<-vegan::ordiArrowMul(arrowmat_nifH_rda)

#Multiply biplot by scaling multiplier
arrowmat_nifH_rda_scale<-arrowmat_nifH_rda*700

# Add labels, make a data.frame
arrowdf_nifH_rda <- data.frame(labels = rownames(arrowmat_nifH_rda_scale), arrowmat_nifH_rda_scale)

# Define the arrow aesthetic mapping
arrow_map <- aes(xend = RDA1,
  yend = RDA2,
  x = 0,
  y = 0,
  shape = NULL,
  color = NULL,
  label = labels)

label_map <- aes(x = 1.3* RDA1,
  y = 1.3 * RDA2,
  shape = NULL,
  color = NULL,
  label = labels)

arrowhead = arrow(length = unit(0.02, "npc"))

# Make a new graphic
rda_plot_nifH +
  geom_segment(
    mapping = arrow_map,
    size = .5,
    data = arrowdf_nifH_rda,
    color = "gray",
    arrow = arrowhead

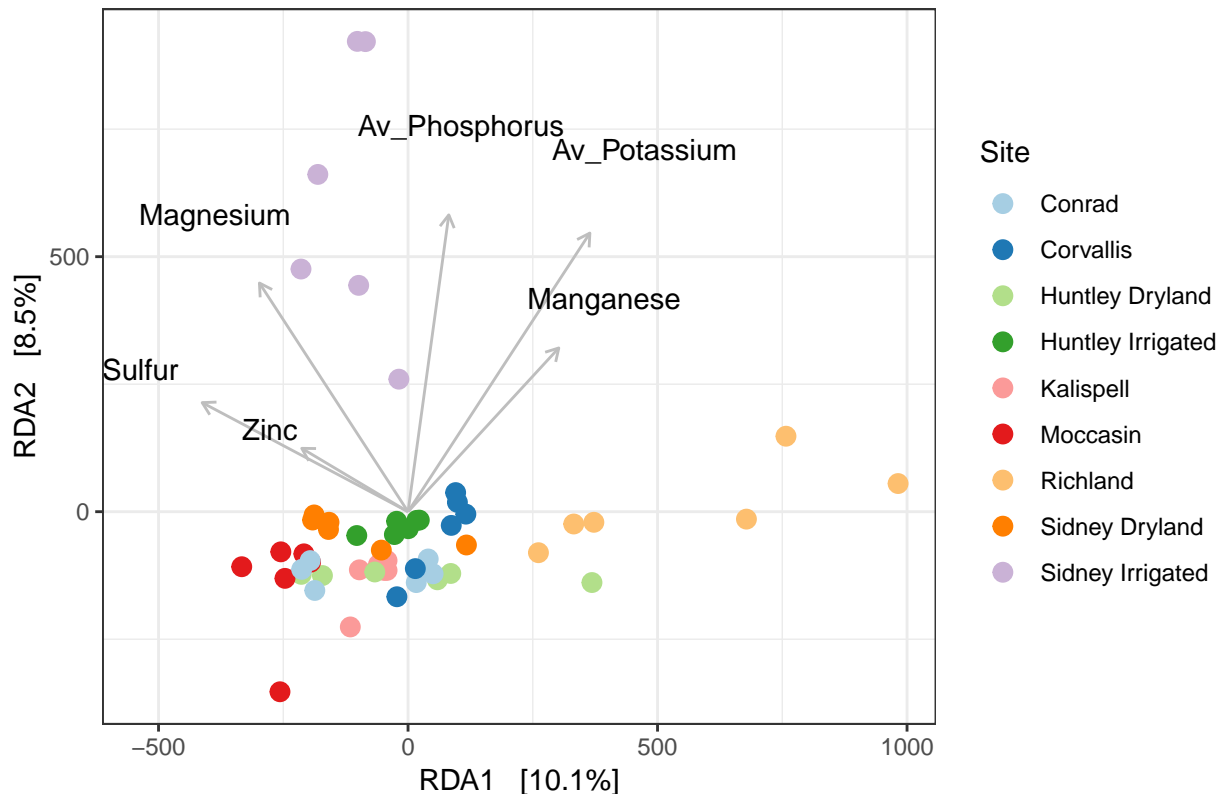
```

```

) +
  geom_text(
    mapping = label_map,
    size = 4,
    data = arrowdf_nifH_rda,
    show.legend = FALSE
  ) +
  ggtitle("RDA plot constrained ordination of nifH with selected Chemistry Model") +
  theme_bw()

```

RDA plot constrained ordination of nifH with selected Chemistry Model



```

# CAP ordinate
cap_ord_nifH <- ordinate(
  physeq = physeq_nifH_ord,
  method = "CAP",
  distance = abund_distnifH,
  formula = ~ Organic_Matter + Barium + Av_Potassium +
    Manganese + Av_Phosphorus + Cobalt + Nitrate_Nitrite)

# CAP plot
cap_plot_nifH <- plot_ordination(
  physeq = physeq_nifH_ord,
  ordination = cap_ord_nifH,
  color = "Site",
  shape = "Plot",
  axes = c(1,2)) +
  geom_point(aes(colour = Site), size = 3) +

```

```

    scale_color_manual(values = farm_col_paired)

# Now add the environmental variables as arrows
arrowmat_nifH_cap <- vegan::scores(cap_ord_nifH, display = "bp")

#Get appropriate scaling multiplier
mul<-vegan::ordiArrowMul(arrowmat_nifH_cap)

#Multiply biplot by scaling multiplier
arrowmat_nifH_cap_scale<-arrowmat_nifH_cap*1.9

# Add labels, make a data.frame
arrowdf_nifH_cap <- data.frame(labels = rownames(arrowmat_nifH_cap_scale), arrowmat_nifH_cap_scale)

# Define the arrow aesthetic mapping
arrow_map <- aes(xend = CAP1,
  yend = CAP2,
  x = 0,
  y = 0,
  shape = NULL,
  color = NULL,
  label = labels)

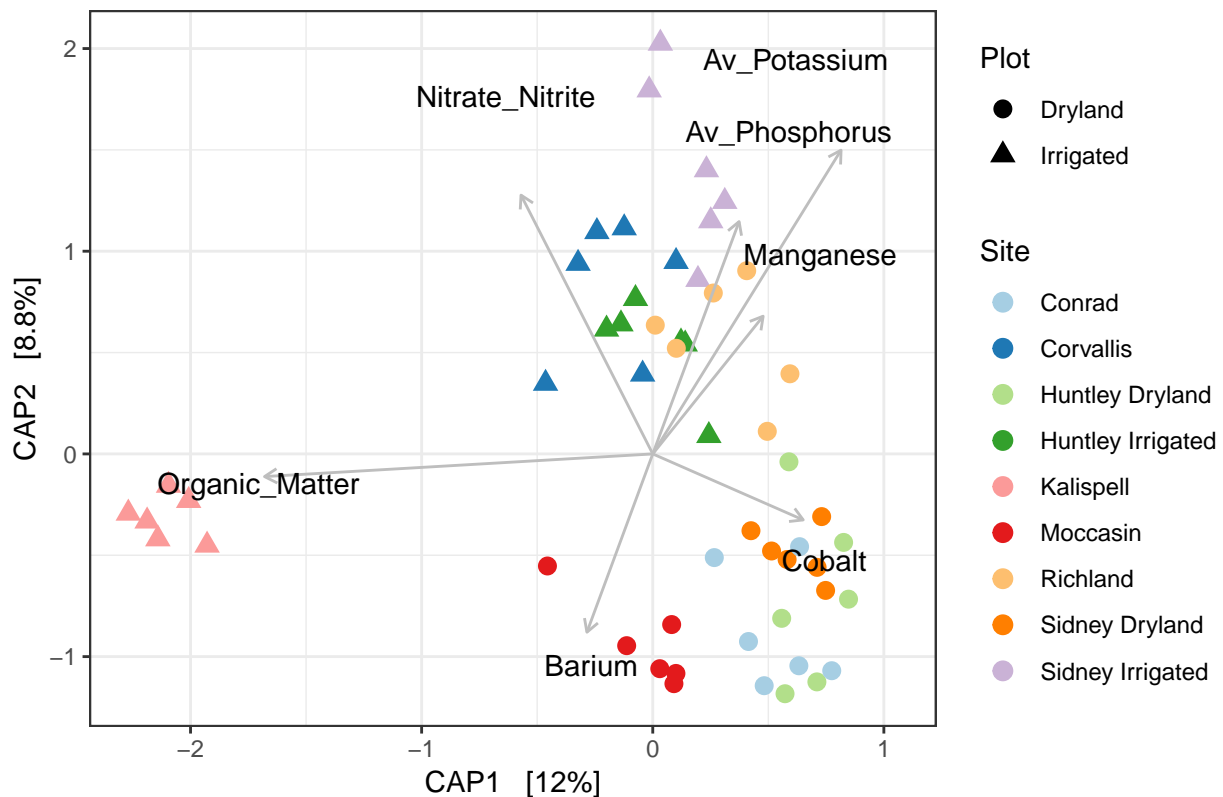
label_map <- aes(x = 1.3 * CAP1,
  y = 1.3 * CAP2,
  shape = NULL,
  color = NULL,
  label = labels)

arrowhead = arrow(length = unit(0.02, "npc"))

# Make a new graphic
cap_plot_nifH +
  geom_segment(
    mapping = arrow_map,
    size = .5,
    data = arrowdf_nifH_cap,
    color = "gray",
    arrow = arrowhead
  ) +
  geom_text_repel(
    mapping = label_map,
    size = 4,
    data = arrowdf_nifH_cap,
    show.legend = FALSE
  ) +
  ggtitle("CAP Plot Constrained Ordination of nifH with Selected Model")+
  theme_bw()

```

CAP Plot Constrained Ordination of nifH with Selected Model



```
{r echo=FALSE} tiff("nifHCAP_chem.tiff", width = 8, height =
5, units = 'in', res = 600) cap_plot_nifH + geom_segment(
mapping = arrow_map, size = .5, data = arrowdf_nifH_cap,
color = "gray", arrow = arrowhead ) + geom_text_repel(
mapping = label_map, size = 4, data = arrowdf_nifH_cap,
show.legend = FALSE )+ ggtitle("CAP Plot Constrained Ordination
of nifH with Selected Model")+ theme_bw() dev.off() #
```

Fitting species to cap plot

```
m1_nifH_cap_chem_species<- capscale(abund_tablenifH~ Organic_Matter + Barium + Av_Potassium +
Manganese + Av_Phosphorus + Cobalt + Nitrate_Nitrite, data = meta3, distance = "bray")
dims=c(1,2)
site=scores(m1_nifH_cap_chem_species,display="wa",choices=dims)
cor.min=0.9 #below this threshold, arrows will be not plotted
#because orrelation is considered too much week
cor_sp=as.data.frame(scores(m1_nifH_cap_chem_species, dis="sp", scaling=1,choices=dims))
cor_sp$cor=with(cor_sp,sqrt(CAP1^2+CAP2^2))
cor_sp$sup=FALSE;cor_sp$sup[cor_sp$cor>=cor.min]<-TRUE
cor_sp$labels=row.names(cor_sp)
cor_sp=cor_sp[cor_sp$sup==TRUE,]
cor_sp_s1=cor_sp
```

```
tax_nifH_cap<- read.delim("~/Alex Alleman/Statewide Microbiome Analysis/Statewide analysis/nifH_OTU_ids,
head(tax_nifH_cap)[,1:8]
```

	kingdom	phylum	class	order	family	genus
OTU1	k__bacteria	p__firmicutes	c__bacilli	o__bacillales	f__paenibacillaceae	g__pae
OTU2	k__bacteria	p__firmicutes	c__bacilli	o__bacillales	f__paenibacillaceae	g__pae
OTU3	k__bacteria	p__proteobacteria	c__alphaproteobacteria	o__rhizobiales	f__rhizobiaceae	g__rhiz
OTU4	k__bacteria	p__firmicutes	c__bacilli	o__bacillales	f__paenibacillaceae	g__pae
OTU5	k__bacteria	p__proteobacteria	c__alphaproteobacteria	o__rhizobiales	f__bradyrhizobiaceae	g__bra
OTU6	k__bacteria	p__proteobacteria	c__alphaproteobacteria	o__rhizobiales	f__bradyrhizobiaceae	g__rho

```
#use perl script to remove teh g_
tax_nifH_cap$species<- sub(".*_", "", tax_nifH_cap$species)
#use the function captitalize to captitalize first letter
tax_nifH_cap$species<- capitalize(tax_nifH_cap$species)

species_nifH<-merge(cor_sp_s1, tax_nifH_cap, by="row.names")
```

Replot with species

```
# CAP ordinate
cap_ord_nifH <- ordinate(
  physeq = physeq_nifH_ord,
  method = "CAP",
  distance = abund_distnifH,
  formula = ~ Organic_Matter + Barium + Av_Potassium +
    Manganese + Av_Phosphorus + Cobalt + Nitrate_Nitrite)

# CAP plot
cap_plot_nifH <- plot_ordination(
  physeq = physeq_nifH_ord,
  ordination = cap_ord_nifH,
  color = "Site",
  shape = "Plot",
  axes = c(1,2)) +
  geom_point(aes(colour = Site), size = 3) +
  scale_color_manual(values = farm_col_paired)

# Define the arrow aesthetic mapping
arrow_map <- aes(xend = CAP1,
  yend = CAP2,
  x = 0,
  y = 0,
  shape = NULL,
  color = NULL,
  label = species)

label_map <- aes(x = 1.3 * CAP1,
  y = 1.3 * CAP2,
  shape = NULL,
```

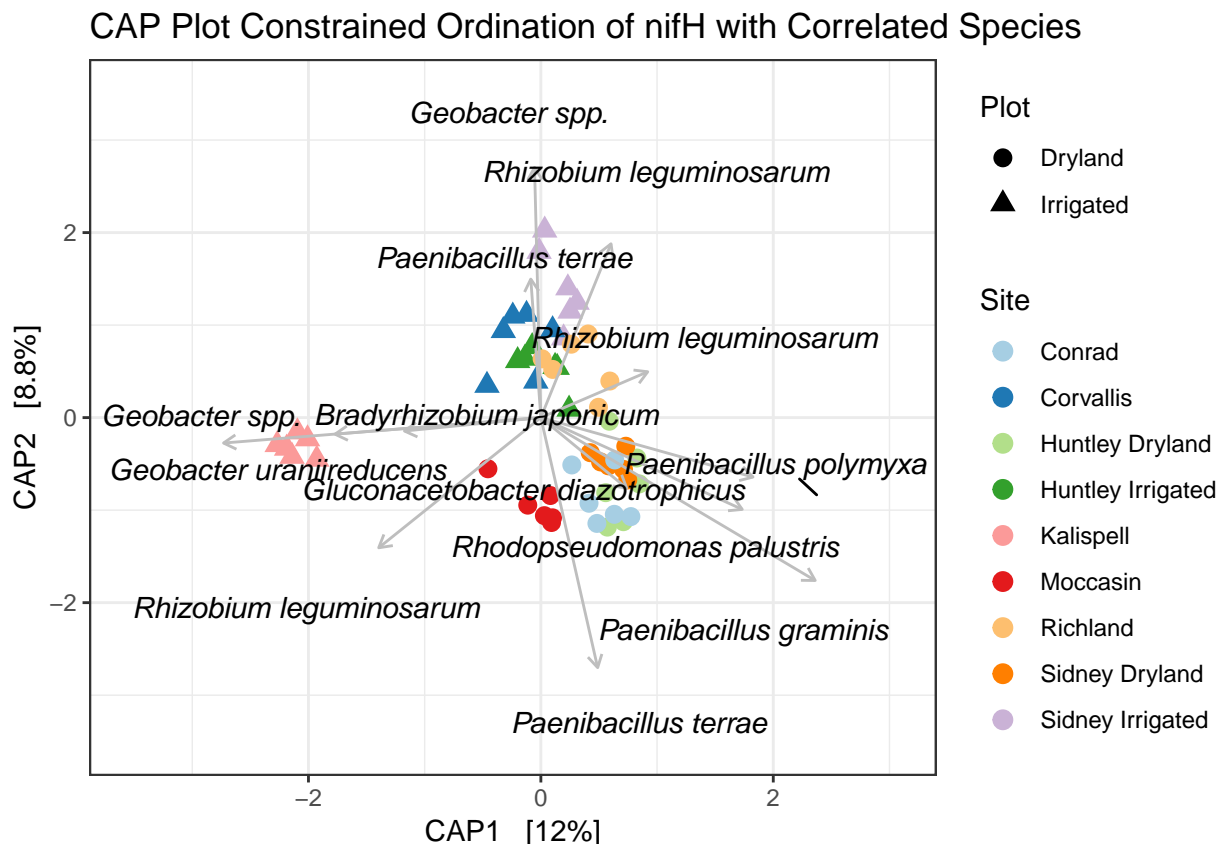
```

color = NULL,
label = species)

arrowhead = arrow(length = unit(0.02, "npc"))

# Make a new graphic
cap_plot_nifH +
  geom_segment(
    mapping = arrow_map,
    size = .5,
    data = species_nifH,
    color = "gray",
    arrow = arrowhead
  ) +
  geom_text_repel(
    mapping = label_map,
    size = 4,
    data = species_nifH,
    show.legend = FALSE,
    fontface="italic"
  ) +
  ggtitle("CAP Plot Constrained Ordination of nifH with Correlated Species") +
  theme_bw()

```



```

Publish to tiff #{r echo=FALSE} tiff("nifHCAP_species.tiff", width = 8, height = 5, units
= 'in', res = 600) cap_plot_nifH +   geom_segment(   mapping = arrow_map,   size =
.5,   data = species_nifH,   color = "gray",   arrow = arrowhead   ) +   geom_text_repel(

```

```
mapping = label_map,      size = 4,      data = species_nifH,      show.legend = FALSE,  
fontface="italic" )+      ggtitle("CAP Plot Constrained Ordination of nifH with Correlated  
Species")+      theme_bw() dev.off() #
```