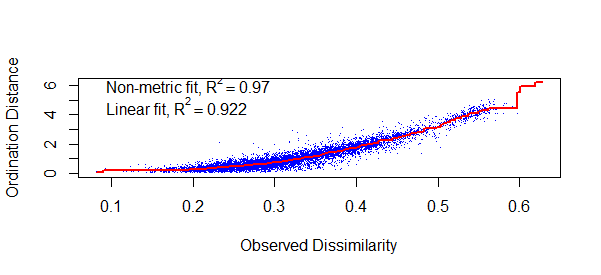
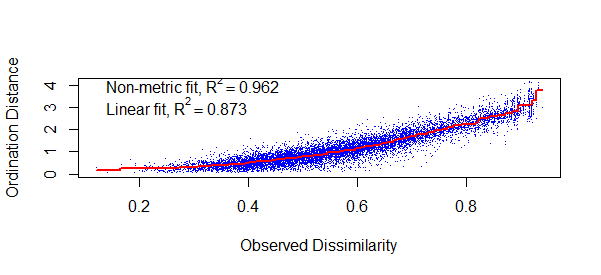
Using Percent Cover I ran metaMDS() using Euclidean distance. This produced a stress of .17 with the following plot:



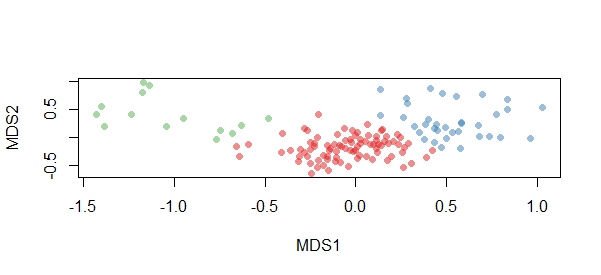
I used Euclidean instead of Bray-Curtis because: 1. K-means uses Euclidean and I am unsure whether it matters if I use two different metrics, and 2. Bray-Curtis produced a stress of .19 with the following stressplot:



Next I decided to use 3 clusters for K-means using the following plot



Using 3 clusters (legend in progress)



Cluster Number 3 2 1

Number of Points in Cluster 14 35 87

# First cluster

[1] "1" "2" "10" "12" "14" "15" "17" "18"

[9] "20" "33" "39" "50" "61" "IL4\_8"

# Second Cluster

[1] "11" "B1\_2" "B1\_3" "B1\_4" "B1\_8" "B2\_1" "B2\_8"

[8] "B3\_3" "B3\_4" "B3\_5" "B3\_9" "B4\_7" "B4\_8" "IL1\_10"

[15] "IL1\_3" "IL1\_8" "IL2\_2" "IL2\_5" "M1\_1" "M1\_10" "M1\_2"

[22] "M1\_4" "M1\_6" "M1\_7" "M1\_8" "M2\_10" "M2\_5" "M2\_6"

[29] "M2\_7" "M2\_8" "M2\_9" "M3\_1" "M3\_7" "M4\_4" "M4\_5"

# Third Cluster

[1] "16" "19" "21" "23" "24" "32" "38"

[8] "40" "42" "43" "44" "47" "48" "57"

[15] "59" "60" "67" "68" "73" "77" "80"

[22] "82" "90" "B1\_1" "B1\_10" "B1\_5" "B1\_6" "B1\_7"

[29] "B1\_9" "B2\_10" "B2\_2" "B2\_4" "B2\_5" "B2\_6" "B2\_7"

[36] "B2\_9" "B3\_1" "B3\_10" "B3\_2" "B3\_6" "B3\_7" "B3\_8"

[43] "B4\_1" "B4\_10" "B4\_2" "B4\_3" "B4\_4" "B4\_6" "B4\_9"

[50] "IL1\_1" "IL1\_4" "IL1\_5" "IL1\_9" "IL2\_1" "IL2\_10" "IL2\_3"

[57] "IL2\_4" "IL2\_7" "IL2\_8" "IL2\_9" "IL3\_1" "IL3\_10" "IL3\_2"

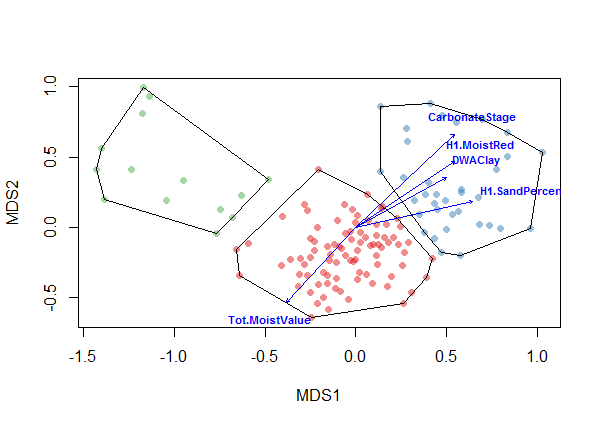
[64] "IL3\_3" "IL3\_4" "IL3\_6" "IL3\_7" "IL3\_9" "IL4\_10" "IL4\_2"

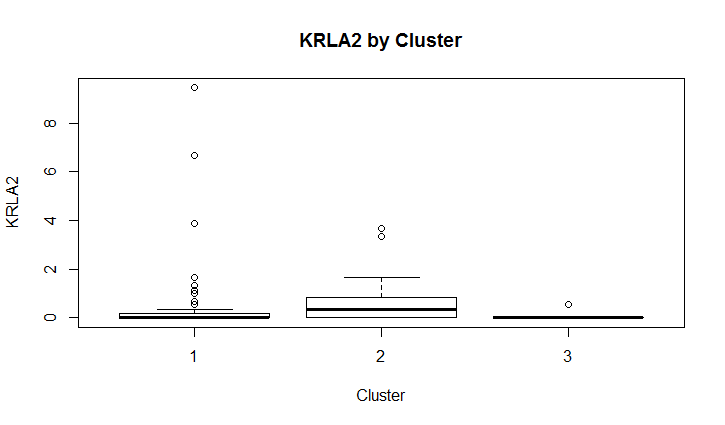
[71] "IL4\_3" "IL4\_4" "IL4\_5" "IL4\_7" "IL4\_9" "M1\_5" "M2\_1"

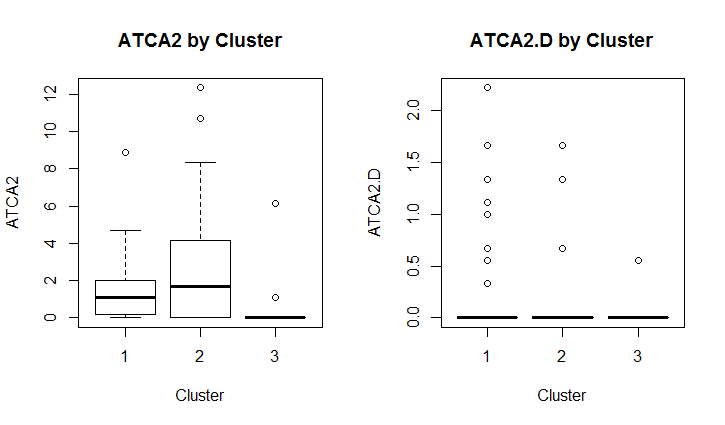
[78] "M2\_3" "M3\_2" "M3\_3" "M3\_5" "M3\_8" "M4\_2" "M4\_3"

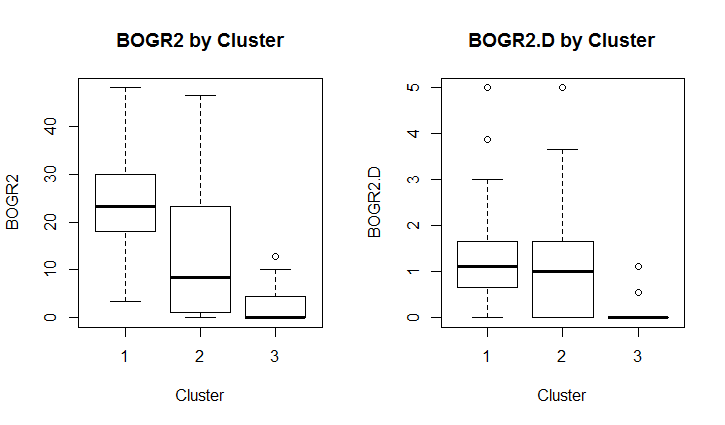
[85] "M4\_6" "M4\_7" "M4\_9"

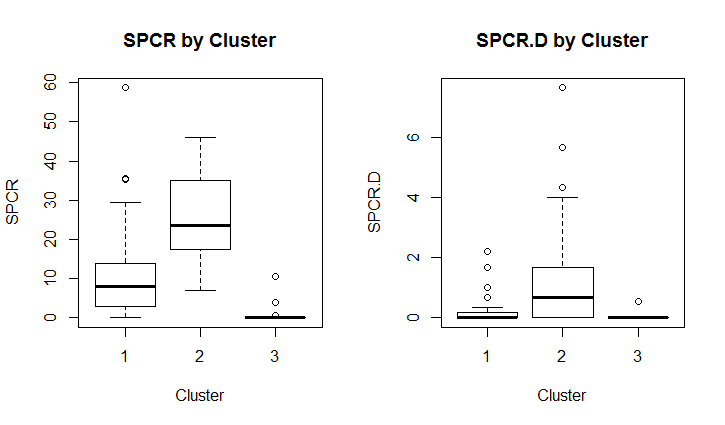
MDS followed by K-Means with environmental soil variables and ordination hulls overlaid

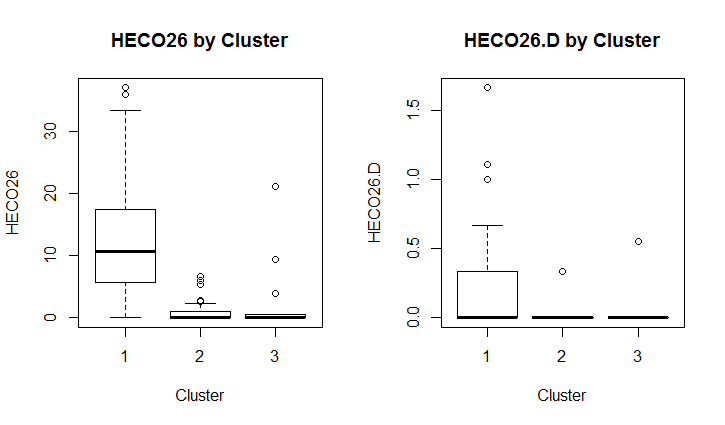


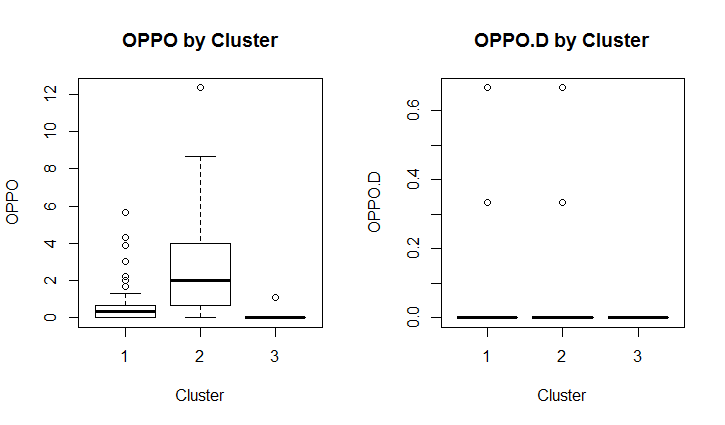


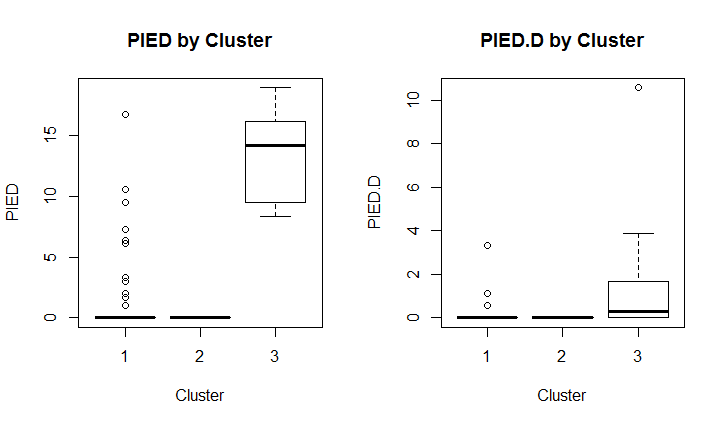


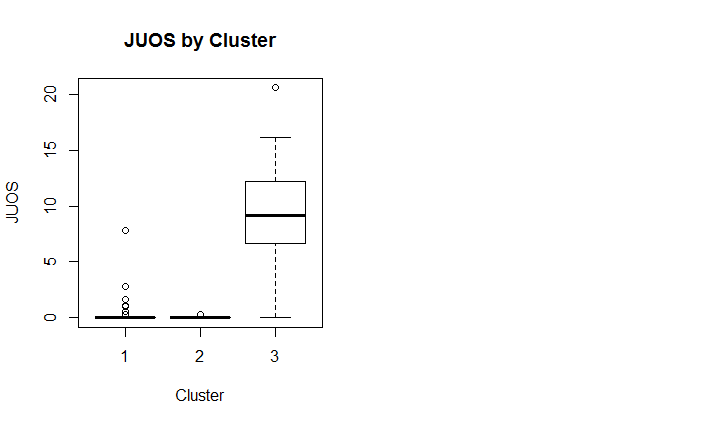




****

****





**R CODE**

# MDS before K-Means 05/11/2016

# LPI Data

data <- read.csv("F:/LPI/Output/USGSLPIPercentCover.csv",header=TRUE, row.names=1)

# plot variance of columns

mar <- par()$mar

par(mar=mar+c(0,5,0,0))

# par(mfrow=c(1,2))

barplot(sapply(data, var), horiz=T, las=1, cex.names=0.5)

barplot(sapply(data, var), horiz=T, las=1, cex.names=0.5, log='x')

par(mar=mar)

######## NMDS #########

library(vegan)

library(MASS)

library(colorspace)

ord<-metaMDS(comm=data,distance="bray",trace=FALSE)

ord #.17

plot.sc = scores(ord)

# Stress <0.10 indicates that the ordination is good "with no real

# risk of drawing false inferences" (Clarke 1993, p. 26).

# linear fit is the fit between ordination values and distances

stressplot(ord)

### K-Means ###

# Determine number of clusters

wss <- (nrow(ord$points)-1)\*sum(apply(ord$points,2,var))

for (i in 2:15) wss[i] <- sum(kmeans(ord$points,

centers=i)$withinss)

plot(1:15, wss, type="b", xlab="Number of Clusters",

ylab="Within groups sum of squares")

# From scree plot elbow occurs at k = 3 (or 8?)

# Apply k-means with k=3 (then try 8)

k <- kmeans(ord$points, 3, nstart=25, iter.max=1000)

library(RColorBrewer)

library(scales)

palette(alpha(brewer.pal(9,'Set1'), 0.5))

plot(ord$points, col=k$clust, pch=16)

legend("topright",c("Cluster 1", "Cluster 2","Cluster 3")

,pch=16, col=k$cluster)

# Cluster sizes

sort(table(k$clust))

clust <- names(sort(table(k$clust)))

# First cluster

row.names(data[k$clust==clust[1],])

# Second Cluster

row.names(data[k$clust==clust[2],])

# Third Cluster

row.names(data[k$clust==clust[3],])

# Compare accommodation by cluster in boxplot

boxplot(data$KRLA2 ~ k$cluster,

xlab='Cluster', ylab='KRLA2',

main='KRLA2 by Cluster')

par(mfrow=c(1,2))

boxplot(data$ARTR2 ~ k$cluster,

xlab='Cluster', ylab='ARTR2',

main='ARTR2 by Cluster')

boxplot(data$ARTR2.D ~ k$cluster,

xlab='Cluster', ylab='ARTR2.D',

main='ARTR2.D by Cluster')

boxplot(data$ATCA2 ~ k$cluster,

xlab='Cluster', ylab='ATCA2',

main='ATCA2 by Cluster')

boxplot(data$ATCA2.D ~ k$cluster,

xlab='Cluster', ylab='ATCA2.D',

main='ATCA2.D by Cluster')

boxplot(data$BOGR2 ~ k$cluster,

xlab='Cluster', ylab='BOGR2',

main='BOGR2 by Cluster')

boxplot(data$BOGR2.D ~ k$cluster,

xlab='Cluster', ylab='BOGR2.D',

main='BOGR2.D by Cluster')

boxplot(data$SPCR ~ k$cluster,

xlab='Cluster', ylab='SPCR',

main='SPCR by Cluster')

boxplot(data$SPCR.D ~ k$cluster,

xlab='Cluster', ylab='SPCR.D',

main='SPCR.D by Cluster')

boxplot(data$HECO26 ~ k$cluster,

xlab='Cluster', ylab='HECO26',

main='HECO26 by Cluster')

boxplot(data$HECO26.D ~ k$cluster,

xlab='Cluster', ylab='HECO26.D',

main='HECO26.D by Cluster')

boxplot(data$OPPO ~ k$cluster,

xlab='Cluster', ylab='OPPO',

main='OPPO by Cluster')

boxplot(data$OPPO.D ~ k$cluster,

xlab='Cluster', ylab='OPPO.D',

main='OPPO.D by Cluster')

boxplot(data$PIED ~ k$cluster,

xlab='Cluster', ylab='PIED',

main='PIED by Cluster')

boxplot(data$PIED.D ~ k$cluster,

xlab='Cluster', ylab='PIED.D',

main='PIED.D by Cluster')

boxplot(data$JUOS ~ k$cluster,

xlab='Cluster', ylab='JUOS',

main='JUOS by Cluster')

### Add in Soils Varibles ###

data.env <- read.csv("F:/Soils/SoilEnvironmentaldataUSGSApril.csv",header=TRUE, row.names=1)

data.env[is.na(data.env)] <- 0 # replace NA with 0

fit.env <- envfit(ord,data.env,perm=1000)

fit.env

### Choose only the significant environmental data

sig.data.env<-data.env[,c(4,11,14,24,27)]

sig.fit.env<-envfit(ord,sig.data.env,perm=1000)

sig.fit.env # Check that you pulled up the right factors.

### Plot Soils and Ordihull ###

plot(ord$points, col=k$clust, pch=16)

legend("topright",c("Cluster 1", "Cluster 2","Cluster 3")

,pch=16, col=k$cluster)

plot(sig.fit.env,col="blue", cex=0.7,font=2)

ordihull(ord, groups = k$clust, display = "sites")