Jordy Taylor title: "Dune Multivariate HW_JNT" output: pdf_document

---

1. NMDS Plot of Dune Plants to Moisture

```
library(vegan)
```

```
## Loading required package: permute

## Loading required package: lattice

## This is vegan 2.3-3
```

```
library(permute)
library(lattice)
data(dune)
data(dune.env)
?dune

dune_mds = metaMDS(dune, trymax = 5)
```

```
## Run 0 stress 0.1192678
## Run 1 stress 0.1183186
## ... New best solution
## ... procrustes: rmse 0.02026936  max resid 0.06495081
## Run 2 stress 0.1183186
## ... procrustes: rmse 0.0001042381  max resid 0.0003551443
## *** Solution reached
```

```
pdf('nmds_dune.pdf')
plot(dune_mds, type='n')
text(dune_mds, 'sp', cex=.6)
# generate vector of colors
color_vect = rev(terrain.colors(6))[-1]
points(dune_mds, 'sites', pch=19,
       col=color_vect[dune.env$Moisture])
legend('topright', paste("Moisture =", 1:5, sep=''),
       col=color_vect, pch=19)



dev.off()
```

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

```r
contrasts(dune.env$Moisture)
```

```
##              .L   .Q         .C
## [1,] -0.6708204  0.5 -0.2236068
## [2,] -0.2236068 -0.5  0.6708204
## [3,]  0.2236068 -0.5 -0.6708204
## [4,]  0.6708204  0.5  0.2236068
```

Describe how you interpret the graphic.

I can see how the various types of plants have a cover response to different levels of moisture. The plants that are scattered around the pink dots, low moisture, but are positioned higher on the Y axis, are more desert like plants- they are able to grow better (have greater cover) in low-moisture environments. Oppossed to the plants scattered around the pink dots positioned lower on the Y axis, which have a very small coverage as a result of low moisture, these plants need more moisture in order to have a greater level of coverage. Basically they don't do so well in a dry environment. The species scattered around the green dots (high mosture at 4), are those that have good coverage at higher mostures, with the exception of those positioned lower on the Y axis around the green dots, which do not grow well in a wet environment. This would imply that these are desert plants and need a dryer environment in order to have better. Moisture levels (indicated by color) imply that there is more variation in plant cover occurring at extremely dry and at extremely moist climates, oppossed to small levels of variation in plant cover occurring at intermediate levels moist environments. Also the spread of species positioned along NMDS2 and the NMDS1 (x is about -1), I would assume that there is a greater diversity of plants that live along the dune with a moisture level of 1. Overall, plant diversity is greatest at moisture 1, but appears there may be some variation in the amount of moisture at these relatively dry locations, which allows for greater diversity (which we can also see by some outliers: Airaprae, Empenigr„ Hyporadi).

What is the goal of creating such a plot? This is a pretty nice and robust technique to allow us to view quantitative, qualitative and/or categorical data sets and allow us to see relationships among different objects, as those oriented closer together are more similar to one another. This is a useful plot because it allows us to look at a multivariate data set of the response variable and plot it against a predictor variable, without using a unit of measure; rather, it is a visual approach based on a distance matrix. Those further away from each other are more dissimilar, while those ordinated closer together are more similar. It's just a nice quick way when looking at multivariate data sets to quickly see clusters in your data set and which groups (in this case, species) show a similar reaction to a predictor variable. Allows for possible relationships to be investigated later on.

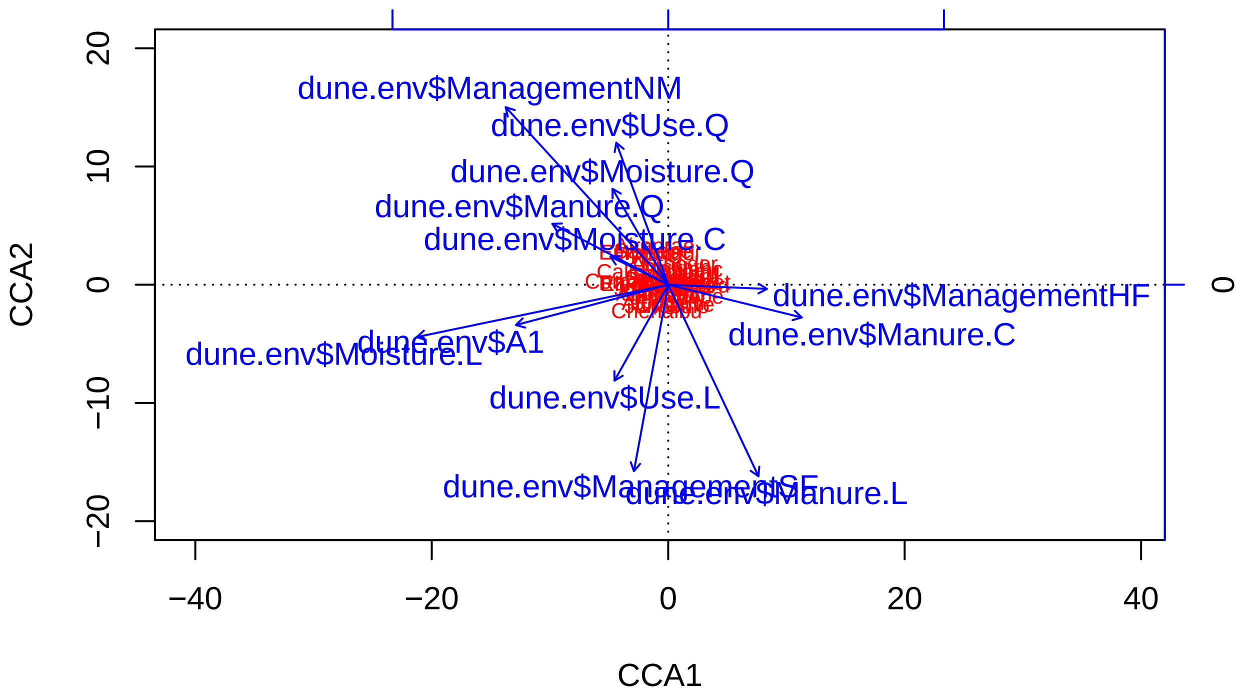
Does this analysis suggest any interesting findings with respect to the dune vegetation?

It appears that more species are oriented around the drier dune (less moisture) areas , implying that there is a greater diversity (forest green) in drier areas. Less species appear to be oriented around the moist areas, implying that there is a lower level of diversity in these areas. Additionally, not only is there a greater diversity of species found at a drier environment, they have a overall greater amounts of coverage than do the species found in moist areas. So there are more species found in a drier environment, and they have greater coverage (on average) than those in a moist environment. It also appears that not many dune areas get intermediate levels of dune vegetation, implying that dunes are predominately extreme conditions with regards to moisture. Not only that, the lowest diversity of plants found out of all the moisture levels is that found at the intermediate level (the ugly yellow green dots). So intermediate moisture is not so good to extremely low or extremely high levels.

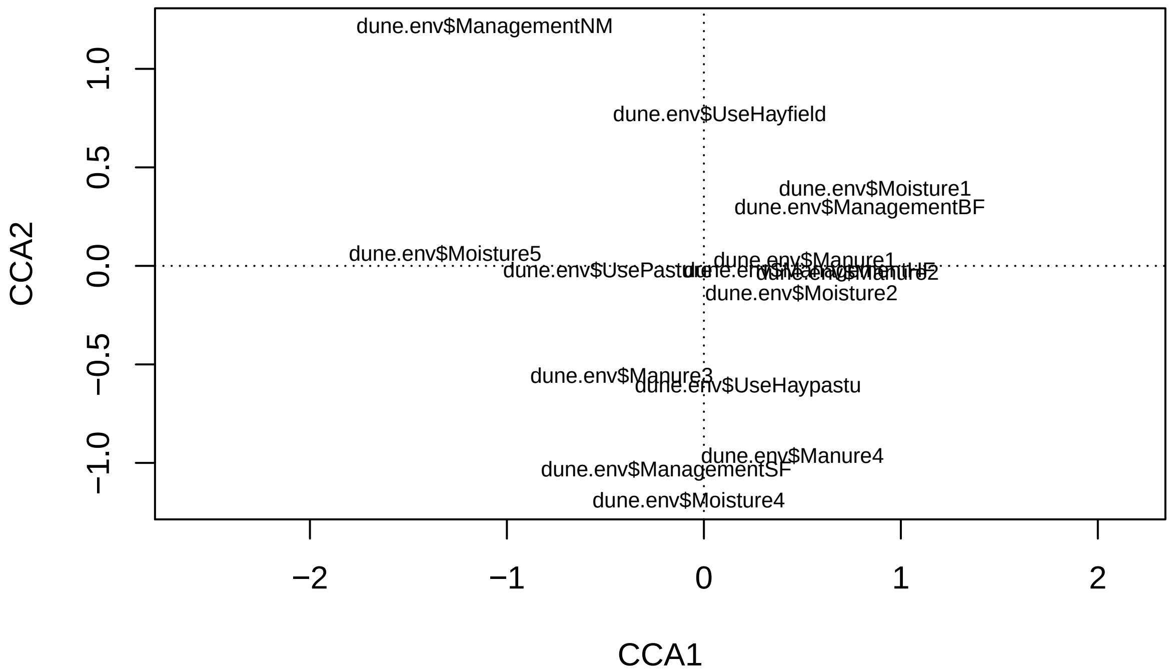
So now I want to look at specific variables that may affect the cover other than moisture. Want to see if all the variables effect or just some.

Carry out a direct ordination using CCA in order to test any potential hypotheses that you developed after examining the MDS plot. Specifically, carry out a test of the entire model (i.e., including all constrained axes) and also carry out tests at the scale of individual explanatory variables you included in your model if you included more than one variable. Plot your results.

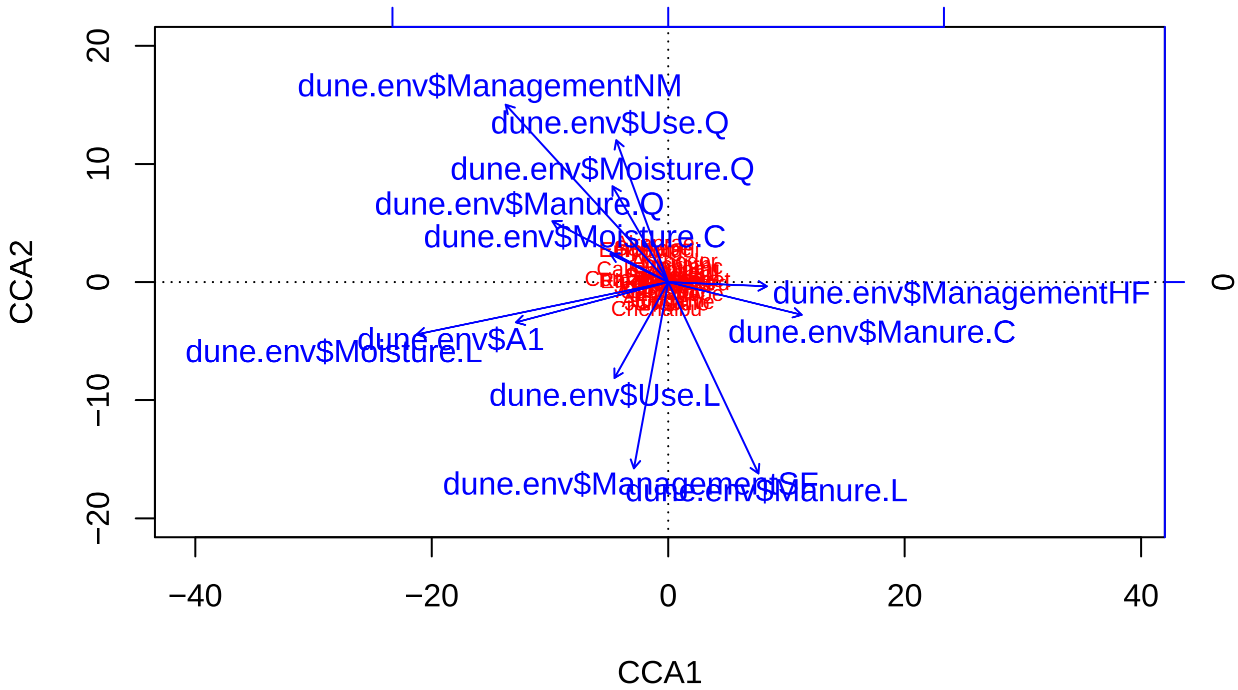
```
#all variables included; all constraints

cca_dune<-cca(dune ~ dune.env$A1 + dune.env$Moisture + dune.env$Management + dune.env$Use + dune.env$Man


#CCA1 and 2 explain a good bit of the variation for the overall data set-all
#constrained
#variables, but the unconstrained CA1 also explains a quarter of the variation.
#So to me, this is questionable. I would say that the unexplained variation
#accounts for a good bit of the variation of the overall model.

plot(cca_dune, ylim=c(-20, 20), display=c('sp','bp'), scaling=1)
```




```
#this lets me view a cca plot of the variables of dune.env
#this is a headache. It appears that since management type go in completely
#different directions than one another, the type of management used might
#have an effect on cover. However, the rest of the variables have quadratic,
#linear and cubic components that explain variation, so this becomes
#extremely difficult to look at , and honestly, I cannot read it.


#So let's try this, instead.

plot(cca_dune, display=c('cn'), scaling=2)
```




```
#this looks a little better. Here we can see that management and manure might
#have a good hold on the coverage since both of the variables are quite dispersed
#from one another and they are found in different quadrants of the plot. Similarily
#with the moisture, being fairly disperesed distance wise from one another on the
#plot (which we previously saw above that has a pretty good effect on cover. so
#here
#we can reassure ourselves that moisture does have a large effect on cover).
#Use does not appear to effect coverage as much, since they are plotted relatively
#close to one another and not in different directions of the plot.


plotdune<-plot(cca_dune, ylim=c(-20, 20), display=c('sp','bp'), scaling=1)
```




```r
#this lets me view the first two cca axes for each species and for each variable

anova(cca_dune, by='margin', permutations = 1000)
```

```
## Permutation test for cca under reduced model
## Marginal effects of terms
## Permutation: free
## Number of permutations: 1000
##
## Model: cca(formula = dune ~ dune.env$A1 + dune.env$Moisture + dune.env$Management + dune.env$Use + du
##                       Df  ChiSquare      F Pr(>F)
## dune.env$A1            1    0.11070 1.2660 0.2418
## dune.env$Moisture      3    0.31587 1.2041 0.2198
## dune.env$Management    2    0.15882 0.9081 0.5534
## dune.env$Use           2    0.13010 0.7439 0.7812
## dune.env$Manure        3    0.25490 0.9717 0.4965
## Residual               7    0.61210
```

```r
anova(cca_dune)
```

```
## Permutation test for cca under reduced model
## Permutation: free
## Number of permutations: 999
##
## Model: cca(formula = dune ~ dune.env$A1 + dune.env$Moisture + dune.env$Management + dune.env$Use + du
##           Df ChiSquare      F Pr(>F)
## Model     12    1.5032 1.4325  0.015 *
## Residual   7    0.6121
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
#the overall model does okay, but there is still a lot of unexplained variance
#occurring
#when we include all of the variables . We can see with the Anova, that moisture
#does a pretty good job of explaining the variation in the data set, which we
#visually saw with 2 different plots. Along with manure, it explains about 25%
#of the variation.
```

Let's look just at Moisture, since we can see that moisture has the largest explained variance associated in the Anova.

```r
cca_dune_moist = cca(dune ~ dune.env$Moisture )
cca_dune_moist
```

```
## Call: cca(formula = dune ~ dune.env$Moisture)
##
##               Inertia Proportion Rank
## Total          2.1153     1.0000
## Constrained    0.6283     0.2970    3
## Unconstrained  1.4870     0.7030   16
## Inertia is mean squared contingency coefficient
##
## Eigenvalues for constrained axes:
##   CCA1   CCA2   CCA3
## 0.4187 0.1330 0.0766
##
## Eigenvalues for unconstrained axes:
##    CA1    CA2    CA3    CA4    CA5    CA6    CA7    CA8    CA9   CA10
## 0.4098 0.2259 0.1761 0.1234 0.1082 0.0908 0.0859 0.0609 0.0566 0.0467
##   CA11   CA12   CA13   CA14   CA15   CA16
## 0.0419 0.0201 0.0143 0.0099 0.0085 0.0080
```

```r
#Here we can see with the CCA1 for the unconstrained axes explain just as much
#of the variation as does the axes for the constrained axes (41%).

anova(cca_dune_moist, by='margin')
```

```
## Permutation test for cca under NA model
## Marginal effects of terms
## Permutation: free
## Number of permutations: 999
##
## Model: cca(formula = dune ~ dune.env$Moisture)
##                   Df ChiSquare      F Pr(>F)
## dune.env$Moisture  3   0.62831 2.2536  0.002 **
## Residual          16   1.48695
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
anova(cca_dune_moist)
```

```
## Permutation test for cca under reduced model
```

```
## Permutation: free
## Number of permutations: 999
##
## Model: cca(formula = dune ~ dune.env$Moisture)
##          Df ChiSquare      F Pr(>F)
## Model     3   0.62831 2.2536  0.001 ***
## Residual 16   1.48695
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

So let's do this again for another variable, let's do manure by itself since it was the next variable that explained the most variation in the overall model.

```
cca_dune_manure = cca(dune ~ dune.env$Manure )
```

our unconstrained eigenvalues account for more of the variation than does our constrained. #Not looking good for a model.

```
anova(cca_dune_manure, by='margin')
```

```
## Permutation test for cca under NA model
## Marginal effects of terms
## Permutation: free
## Number of permutations: 999
##
## Model: cca(formula = dune ~ dune.env$Manure)
##                 Df ChiSquare      F Pr(>F)
## dune.env$Manure  4   0.61156 1.5251  0.022 *
## Residual        15   1.50370
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova(cca_dune_manure)
```

```
## Permutation test for cca under reduced model
## Permutation: free
## Number of permutations: 999
##
## Model: cca(formula = dune ~ dune.env$Manure)
##          Df ChiSquare      F Pr(>F)
## Model     4   0.61156 1.5251  0.021 *
## Residual 15   1.50370
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

“‘

very similar to what happened with moisture. When we just include manure in the model, it explains a lot more variation when it stands alone and shows a #significant effect on cover.

So we can see that when we include all of the variables in the cca and run an anova, moisture explains the most variation out of all of the other variables, however, it's still not much compared to the overall variation in the data set. So let's isolate just moisture to see what happens. And we can see that when we do that, and run an ANOVA, we can see that the amount of variation that moisture explains increases. We can also see that our model, when including only moisture as the predictor, explains the data much better. So basically, the model with only moisture as an explanatory variable is the best fit model. This is because when we include all of the other variables in the model, these are taking away from the degrees of freedom and are taking away from the variance explained. This ultimately reduces the amount of variance explained for moisture when we include all of the other variables.

```r
library(dummies)
```

```
## dummies-1.5.6 provided by Decision Patterns
```

```r
Moisture<-dummy(dune.env$Moisture)
Management<-dummy(dune.env$Management)
Manure<-dummy(dune.env$Manure)
Use<-dummy(dune.env$Use)
A1<-(dune.env$A1)

landuse<- dune.env[ , c('Management', 'Use', 'Manure')]


varpart(dune, Moisture, Manure, Management)
```

```
## Warning: collinearity detected in X1: mm = 4, m = 3
```

```
## Warning: collinearity detected in X2: mm = 5, m = 4
```

```
## Warning: collinearity detected in X3: mm =4, m =3
```

```
## Warning: collinearity detected in cbind(X1,X2): mm = 9, m = 7
```

```
## Warning: collinearity detected in cbind(X1,X3): mm = 8, m = 6
```

```
## Warning: collinearity detected in cbind(X2,X3): mm = 9, m = 6
```

```
## Warning: collinearity detected in cbind(X1,X2,X3): mm = 13, m = 9
```

```
##
## Partition of variation in RDA
##
## Call: varpart(Y = dune, X = Moisture, Manure, Management)
##
## Explanatory tables:
## X1:  Moisture
## X2:  Manure
## X3:  Management
##
## No. of explanatory tables: 3
## Total variation (SS): 1598.4
##             Variance: 84.124
## No. of observations: 20
##
```

```
## Partition table:
##                       Df R.square Adj.R.square Testable
## [a+d+f+g] = X1         3  0.32674      0.20050     TRUE
## [b+d+e+g] = X2         4  0.34255      0.16723     TRUE
## [c+e+f+g] = X3         3  0.34747      0.22512     TRUE
## [a+b+d+e+f+g] = X1+X2  7  0.58439      0.34195     TRUE
## [a+c+d+e+f+g] = X1+X3  6  0.55186      0.34503     TRUE
## [b+c+d+e+f+g] = X2+X3  6  0.50597      0.27795     TRUE
## [a+b+c+d+e+f+g] = All  9  0.66475      0.36303     TRUE
## Individual fractions
## [a] = X1 | X2+X3       3               0.08508     TRUE
## [b] = X2 | X1+X3       3               0.01800     TRUE
## [c] = X3 | X1+X2       2               0.02109     TRUE
## [d]                    0               0.03483    FALSE
## [e]                    0               0.12345    FALSE
## [f]                    0               0.08963    FALSE
## [g]                    0              -0.00904    FALSE
## [h] = Residuals                        0.63697    FALSE
## Controlling 1 table X
## [a+d] = X1 | X3        3               0.11991     TRUE
## [a+f] = X1 | X2        3               0.17472     TRUE
## [b+d] = X2 | X3        3               0.05283     TRUE
## [b+e] = X2 | X1        4               0.14145     TRUE
## [c+e] = X3 | X1        3               0.14453     TRUE
## [c+f] = X3 | X2        2               0.11072     TRUE
## ---
## Use function 'rda' to test significance of fractions of interest

## Warning: collinearity detected: redundant variable(s)  between tables X2, X3
## results are probably incorrect: remove redundant variable(s) and repeat the analysis

## Warning: collinearity detected: redundant variable(s)  between tables X1, X2, X3
## results are probably incorrect: remove redundant variable(s) and repeat the analysis
```

here we can see that there is strong collinearity between manure and management, so these are redundant variables. X1 + X2 explains 34% of the variation as well as x1 + x3 explains the same amount of variation. THus one of these can be excluded because it is redundant in the analysis.

“‘

So let’s run this again without one of these variables: take out Managmenet.

```r
varpart(dune, Moisture, Manure, A1)
```

```
## Warning: collinearity detected in X1: mm = 4, m = 3

## Warning: collinearity detected in X2: mm = 5, m = 4

## Warning: collinearity detected in cbind(X1,X2): mm = 9, m = 7

## Warning: collinearity detected in cbind(X1,X3): mm = 5, m = 4

## Warning: collinearity detected in cbind(X2,X3): mm = 6, m = 5
```

```
## Warning: collinearity detected in cbind(X1,X2,X3): mm = 10, m = 8

##
## Partition of variation in RDA
##
## Call: varpart(Y = dune, X = Moisture, Manure, A1)
##
## Explanatory tables:
## X1:  Moisture
## X2:  Manure
## X3:  A1
##
## No. of explanatory tables: 3
## Total variation (SS): 1598.4
##             Variance: 84.124
## No. of observations: 20
##
## Partition table:
##                       Df R.square Adj.R.square Testable
## [a+d+f+g] = X1         3  0.32674      0.20050     TRUE
## [b+d+e+g] = X2         4  0.34255      0.16723     TRUE
## [c+e+f+g] = X3         1  0.09646      0.04627     TRUE
## [a+b+d+e+f+g] = X1+X2  7  0.58439      0.34195     TRUE
## [a+c+d+e+f+g] = X1+X3  4  0.35382      0.18150     TRUE
## [b+c+d+e+f+g] = X2+X3  5  0.40751      0.19591     TRUE
## [a+b+c+d+e+f+g] = All  8  0.61194      0.32971     TRUE
## Individual fractions
## [a] = X1 | X2+X3       3               0.13381     TRUE
## [b] = X2 | X1+X3       4               0.14821     TRUE
## [c] = X3 | X1+X2       1              -0.01223     TRUE
## [d]                    0               0.00143    FALSE
## [e]                    0              -0.00677    FALSE
## [f]                    0               0.04091    FALSE
## [g]                    0               0.02435    FALSE
## [h] = Residuals                        0.67029    FALSE
## Controlling 1 table X
## [a+d] = X1 | X3        3               0.13524     TRUE
## [a+f] = X1 | X2        3               0.17472     TRUE
## [b+d] = X2 | X3        4               0.14964     TRUE
## [b+e] = X2 | X1        4               0.14145     TRUE
## [c+e] = X3 | X1        1              -0.01900     TRUE
## [c+f] = X3 | X2        1               0.02868     TRUE
## ---
## Use function 'rda' to test significance of fractions of interest
```

here less collinearity detected, and here it does a better job at looking at the dispersion of the variation. However, X3 does not explain as much varation
in this model as much as the other and only increases the variance explained only by a little.

```r
varpart(dune, Moisture, Manure, Use)
```

```
## Warning: collinearity detected in X1: mm = 4, m = 3

## Warning: collinearity detected in X2: mm = 5, m = 4
```

```
## Warning: collinearity detected in X3: mm =3, m =2

## Warning: collinearity detected in cbind(X1,X2): mm = 9, m = 7

## Warning: collinearity detected in cbind(X1,X3): mm = 7, m = 5

## Warning: collinearity detected in cbind(X2,X3): mm = 8, m = 6

## Warning: collinearity detected in cbind(X1,X2,X3): mm = 12, m = 9

##
## Partition of variation in RDA
##
## Call: varpart(Y = dune, X = Moisture, Manure, Use)
##
## Explanatory tables:
## X1:  Moisture
## X2:  Manure
## X3:  Use
##
## No. of explanatory tables: 3
## Total variation (SS): 1598.4
##             Variance: 84.124
## No. of observations: 20
##
## Partition table:
##                      Df R.square Adj.R.square Testable
## [a+d+f+g] = X1        3  0.32674      0.20050     TRUE
## [b+d+e+g] = X2        4  0.34255      0.16723     TRUE
## [c+e+f+g] = X3        2  0.12137      0.01800     TRUE
## [a+b+d+e+f+g] = X1+X2 7  0.58439      0.34195     TRUE
## [a+c+d+e+f+g] = X1+X3 5  0.44058      0.24079     TRUE
## [b+c+d+e+f+g] = X2+X3 6  0.42037      0.15284     TRUE
## [a+b+c+d+e+f+g] = All 9  0.63975      0.31552     TRUE
## Individual fractions
## [a] = X1 | X2+X3      3               0.16268     TRUE
## [b] = X2 | X1+X3      4               0.07474     TRUE
## [c] = X3 | X1+X2      2              -0.02642     TRUE
## [d]                   0               0.06011    FALSE
## [e]                   0               0.06671    FALSE
## [f]                   0               0.01204    FALSE
## [g]                   0              -0.03432    FALSE
## [h] = Residuals                       0.68448    FALSE
## Controlling 1 table X
## [a+d] = X1 | X3       3               0.22279     TRUE
## [a+f] = X1 | X2       3               0.17472     TRUE
## [b+d] = X2 | X3       4               0.13485     TRUE
## [b+e] = X2 | X1       4               0.14145     TRUE
## [c+e] = X3 | X1       2               0.04029     TRUE
## [c+f] = X3 | X2       2              -0.01439     TRUE
## ---
## Use function 'rda' to test significance of fractions of interest
```

again, USE does not explain too much of the variation and only shares a small portion of shared variation with X1 and X2.

```
Overall, I would say that the models generally agree with one another.
Yes, they do all show that moisture does the best job at explaining the
most variation in the overall data set.
I would say that they do compliment each other quite nicely, at looking at
how much overall explained variation can be accounted for and where it is
distributed
amongst the included predictor variables. We can see that most of the variation
is distributed towards the moisture, however when we include other variables,
that variation gets more spread because we are losing our degrees of freedom.
Which concurs with the anova when we include all of the variables, as oppossed
to only including one. Our model becomes a better fit when we exclude variables
that are collinear or redundant. So really, only when we include moisture and
manure in our model, does it really fit the data more truely.
I would trust the CCA plot much better along with out ANOVA because it allows
us to see the change in the amount of variation distribution when we include or
exclude different predictor variables. This is really important in modeling, in
general, because including the variables that are
true explanatory variables of variation will better fit your dataset overall.
I think I prefer the CCA with the anova much better because it tends to
cluster the variables
(of different levels/types as well) into different areas, based on distance and direction.
TO me, it's easier to decipher when deciding
which variables are linked with one another and can be excluded
since they are being redundant.

While this doesn't take into account of the variance
explained by each variable (visually),
we can see which variables explain variation by assessing the eigenvalues.
```