Seminar 7: multivariate

Jon Bray/Olivia Burge

23 September 2014

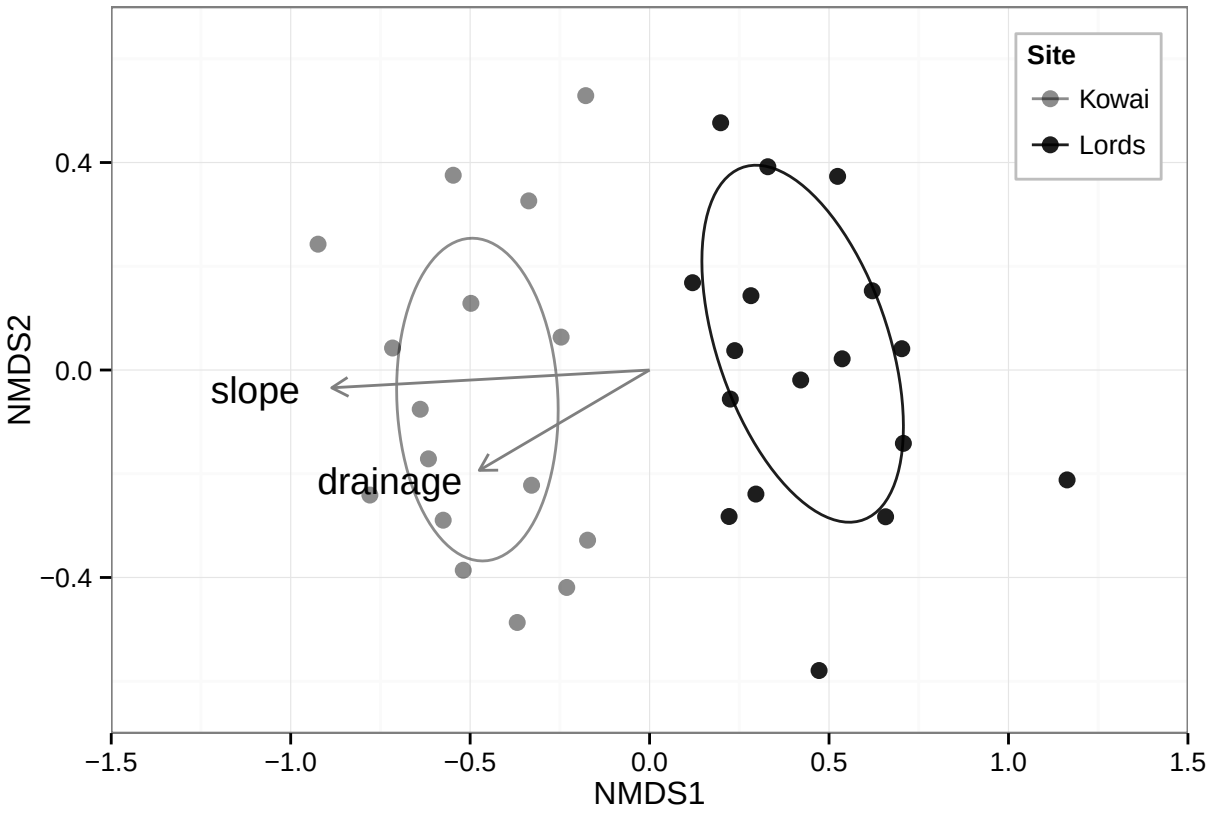
Multivariate data - introduction

We can put a bunch of stuff here on why on earth one ends up with multiple responses, multiple predictors and decide it's a good idea to put it into R.

A common workflow with multivariate data is:

1. Visualise it (plotting)
2. Basic ordination (more on this later)
3. Run more detailed multivariate analyses and ordinations
4. Plot the results and write up

You may remember doing something like this in undergrad:



Ordination methods

[Something about constrained vs unconstrained]

Unconstrained: NMDS and the rest

Before we start our analysis, we need to examine the data. Often we might histogram out all the species to what they are doing, and do some

```
#data - from the vegan package:
#see ?mite for an explanation of what it is
data(mite)
head(mite[, 1:11], 3) #shows us first 3 lines of first 11 columns
```

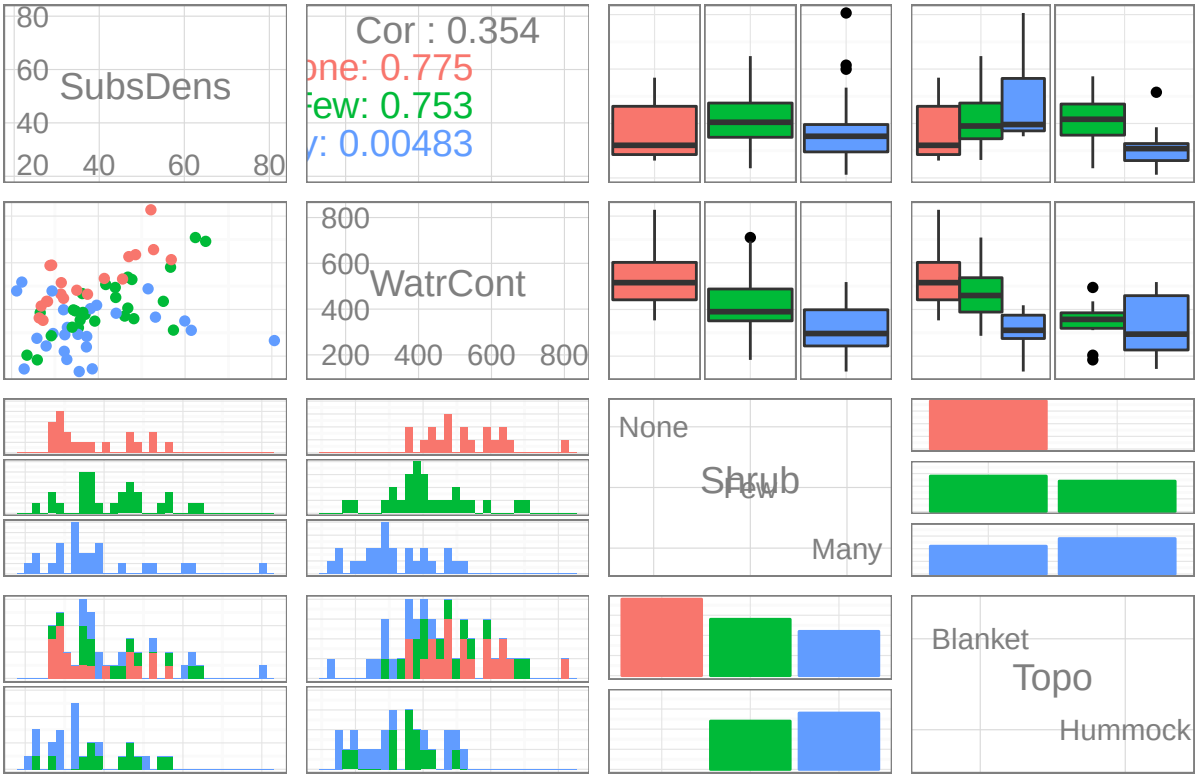
```
##   Brachy PHTH HPAV RARD SSTR Protopl MEGR MPRO TVIE HMIN HMIN2
## 1     17    5    5    3    2       1    4    2    2    1     4
## 2      2    7   16    0    6       0    4    2    0    0     1
## 3      4    3    1    1    2       0    3    0    0    0     6
```

```
data(mite.env)
head(mite.env, 3) #first three lines
```

```
##   SubsDens WatrCont Substrate Shrub    Topo
## 1    39.18    350.1   Sphagn1   Few Hummock
## 2    54.99    434.8    Litter   Few Hummock
## 3    46.07    371.7 Interface   Few Hummock
```

We might also want to do a quick visualisation of the data. Here we leave out Substrate because it's a bit slow.

```
ggpairs(mite.env[c(1:2,4:5)], colour = "Shrub") # yes the colours are atrocious.
```

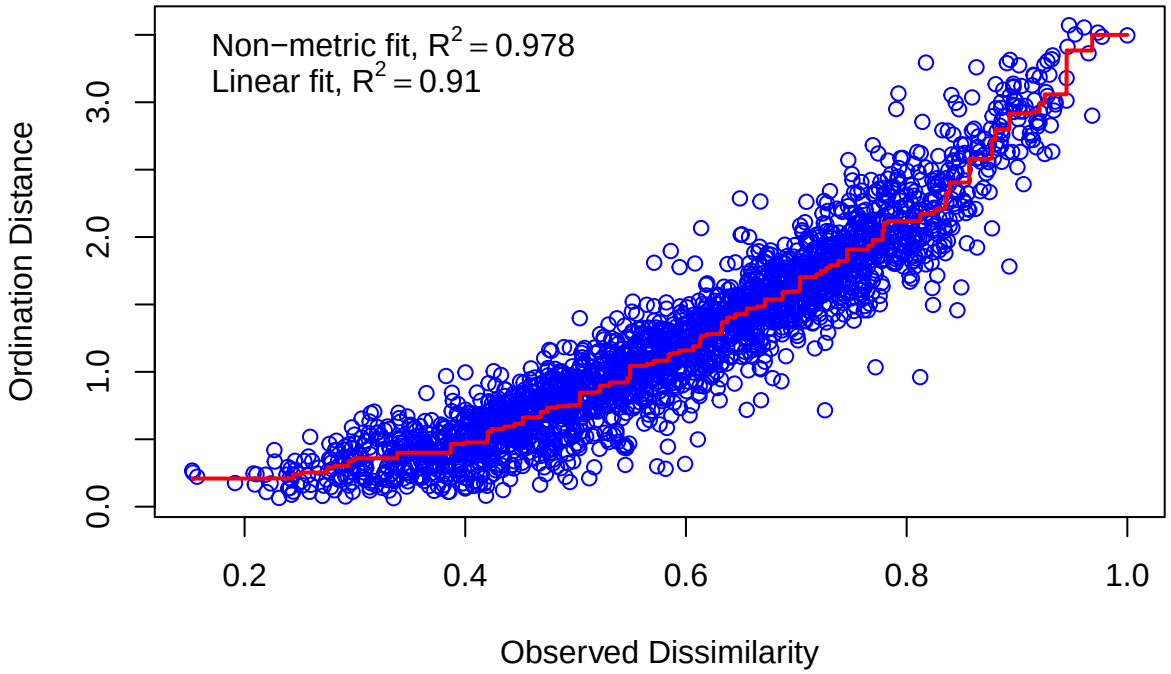


```
# alternatively, use the base version - faster - we can include Substrate, but less user  friendlly
pairs(mite.env)
```

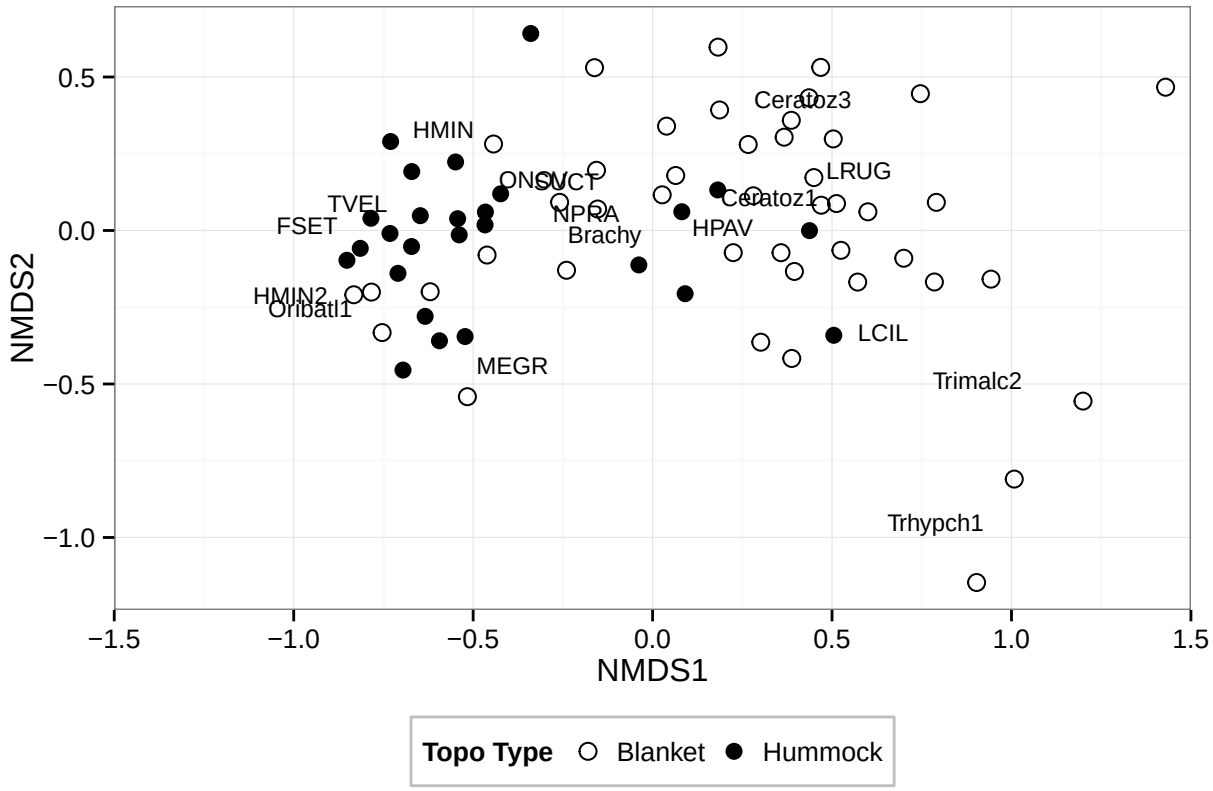
NMDS

```
set.seed(505) #set seed for reproducibility.
meta.nmds.mite <- metaMDS(mite) #no transformation of species data is made here
# prior to bray curtis dissimilarities being calculated.
#(Bray Curtis is the default in R).
meta.nmds.mite #prints very basic dsecription
```

```
stressplot(meta.nmds.mite) # To gain the stress plot for stress values for your NMDS
```



Once we've run the NMDS, we can plot the sites and their species in a plot. Here we've coloured them by topo type out of interest.



Effect of environment on composition (envfit)

We may also wish to see how environmental variables we've measured at each plot correlate with the species composition at each plot. Envfit runs a correlation:

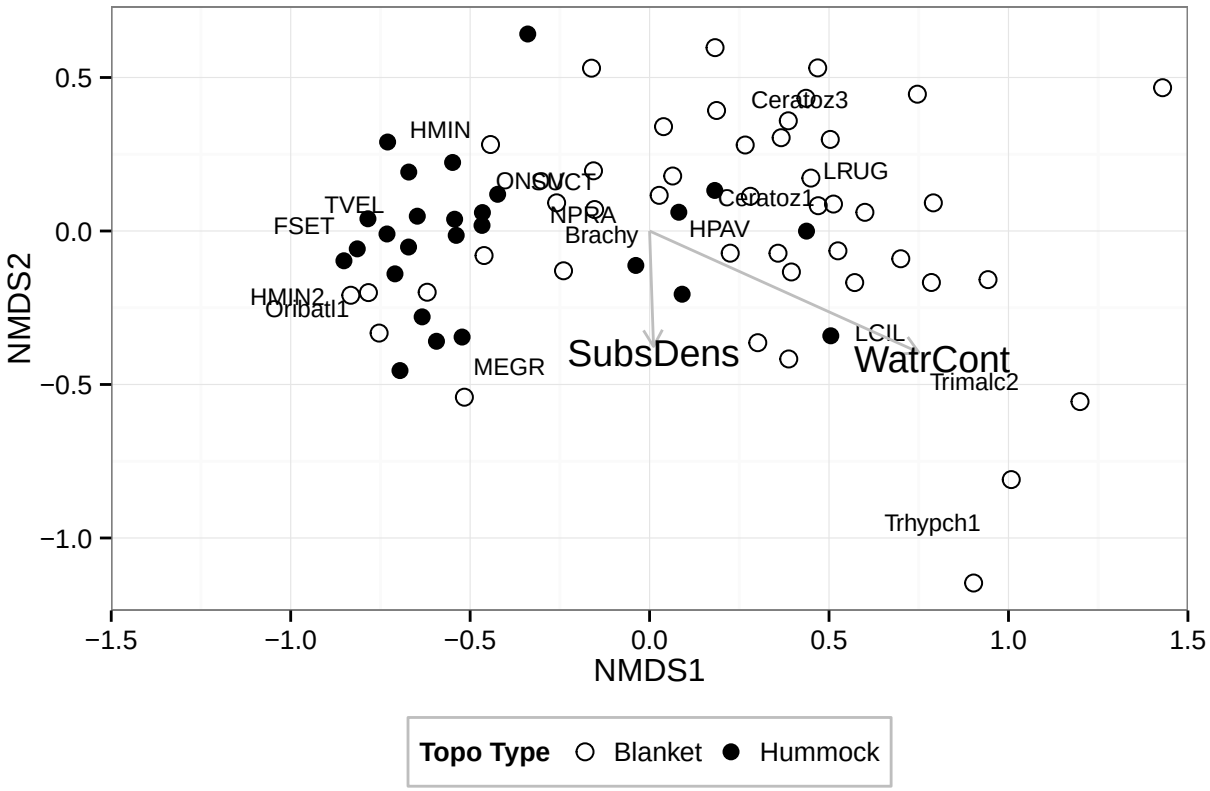
```
mite.envfit <- envfit(meta.nmds.mite, env = mite.env, perm = 999) #standard envfit
mite.envfit
```

```
##
## ***VECTORS
##
##           NMDS1  NMDS2   r2 Pr(>r)
## SubsDens  0.029 -1.000 0.14  0.006 **
## WatrCont  0.885 -0.466 0.70  0.001 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## P values based on 999 permutations.
##
## ***FACTORS:
##
## Centroids:
##                     NMDS1 NMDS2
## SubstrateSphagn1    -0.04 -0.04
## SubstrateSphagn2     0.24  0.10
## SubstrateSphagn3    -0.62 -0.20
## SubstrateSphagn4    -0.44  0.09
## SubstrateLitter     -0.72 -0.39
```

```
## SubstrateBarepeat   1.17 -0.34
## SubstrateInterface -0.04  0.06
## ShrubNone           0.64  0.00
## ShrubFew           -0.07  0.02
## ShrubMany          -0.41 -0.01
## TopoBlanket         0.25  0.01
## TopoHummock        -0.43 -0.02
##
## Goodness of fit:
##             r2 Pr(>r)
## Substrate 0.20  0.006 **
## Shrub     0.40  0.001 ***
## Topo      0.25  0.001 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## P values based on 999 permutations.
```

We see that the two continuous variables, substrate density ("SubsDens") and water content ("WatrCont") have significant effects, particulary water content (r-squared of 0.7). The factors - substrate, shrub, and topo are all significant as well.

We can plot the continuous variables as vectors on the ordination:

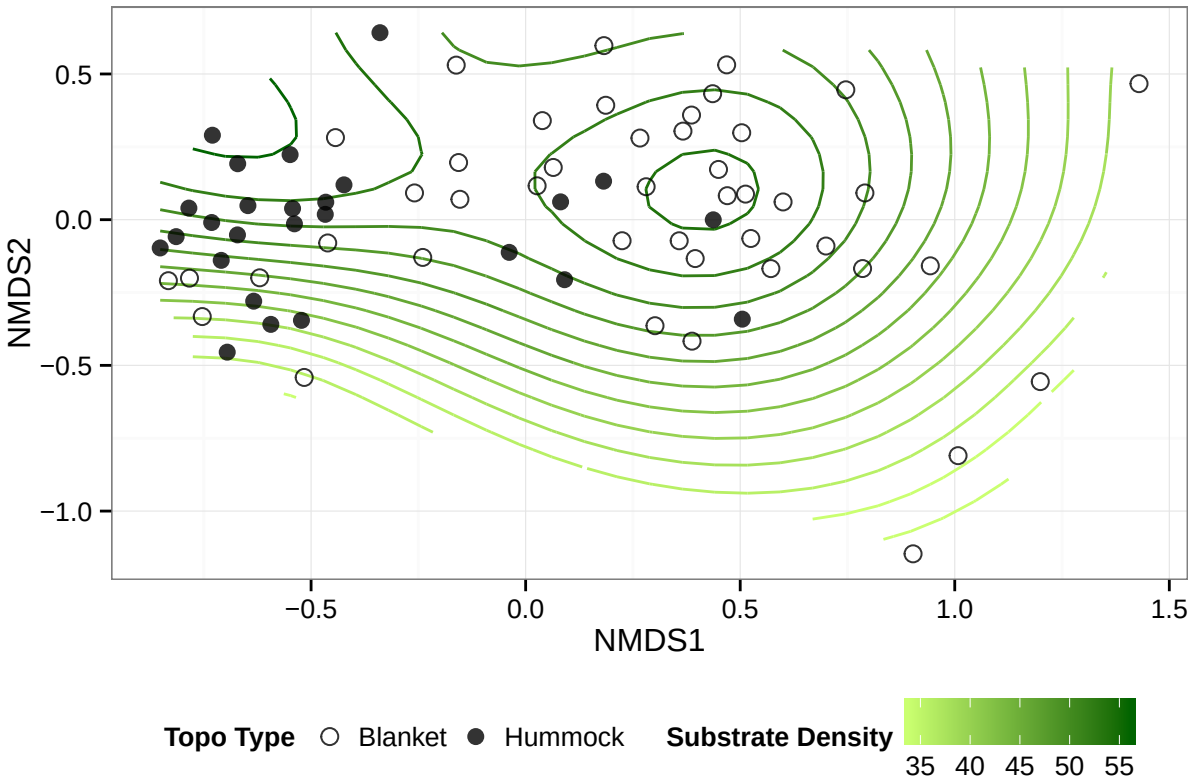


Non-linear environmental effects (ordisurf)

Environmental effects aren't always linear though. We can check for this by using the ordisurf function in vegan (which is even recommended in the envfit() help!):

```
# we use the function ordisurf to look at the relationship to one of the variables with the biggest inf
ordi.mite <- ordisurf(meta.nmds.mite ~ mite.env$SubsDens) #created the ordisurf object
```

We can also plot in ggplot2 (lots of extra code hidden):



Multivariate anova (permanova/adonis)

[Some smart text about the difference between adonis and envfit. I have some basic stuff somewhere]

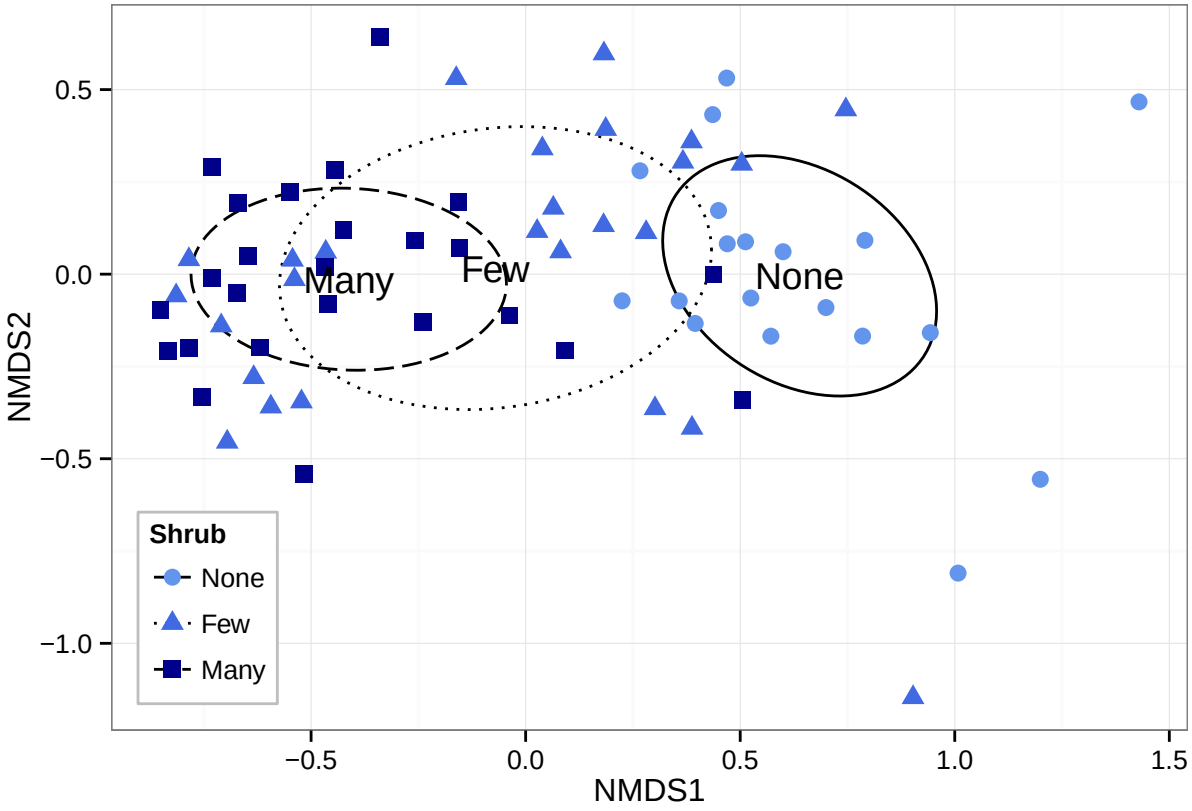
```
#scaling env variables
mite.env$SubsDens_scaled <- scale(mite.env$SubsDens) #scaling the variables
mite.env$WatrCont_scaled <- scale(mite.env$WatrCont)

#running the permanova
adonis.mite <- adonis(formula = mite ~ SubsDens_scaled + WatrCont_scaled + Substrate + Shrub + Topo,
                      data = mite.env,
                      method = "bray")
adonis.mite
```

```
##
## Call:
## adonis(formula = mite ~ SubsDens_scaled + WatrCont_scaled + Substrate +      Shrub + Topo, data = mi
##
## Terms added sequentially (first to last)
##
##                 Df SumsOfSqs MeanSqs F.Model    R2 Pr(>F)
## SubsDens_scaled  1      0.49    0.49    3.92 0.033  0.001 ***
```

```
## WatrCont_scaled  1      3.64    3.64   28.99 0.247  0.001 ***
## Substrate        6      1.72    0.29    2.29 0.117  0.001 ***
## Shrub            2      0.85    0.42    3.38 0.058  0.002 **
## Topo             1      0.73    0.73    5.81 0.050  0.001 ***
## Residuals       58      7.27    0.13         0.495
## Total           69     14.70                 1.000
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We can plot the results of this - for example, we may wish to see how the closely our plots are when divided on shrub type (our ordinal factor):



Are the groups different, or is their variance different? (betadisper)

Whatever anosim is (anosim)

Constrained

How much variance do the variables explain? (RDA, varpart)

Something about stepwise

Incorporating distance: dbRDA