ExaminingCovariatesI

PetraGuy

27 March 2018

Several approaches will be used to examine the colinearity of the variates used in the to explain he richness of the 16 plots in each site, and the response of the richness to the variates will be considered,

A) Visref will be used to plot scatter graphs of individual variables used in a simple additive multiple linear regression. Using visref to look at the variables allows a multiple regression to be carried out and the effect of each variable is shown with the other variables held constant. This allows for the effect of each variable on the richness, in the presence of the other variables to be considered.
B) Pairs plots will be used to examine colinearity between all variables.
C) PCA will be used to see if we can extract a few key variables to represent the data.
D) VIR values will be calculated
E) p values for the variables will be recorded as variables are removed from the regression. F)ANOVA will be used to see if there is a difference between the means

Since there are a lot of variables, they have been split in two, those which directly relate to abiotic features and those which represent the heterogeneity of the site:

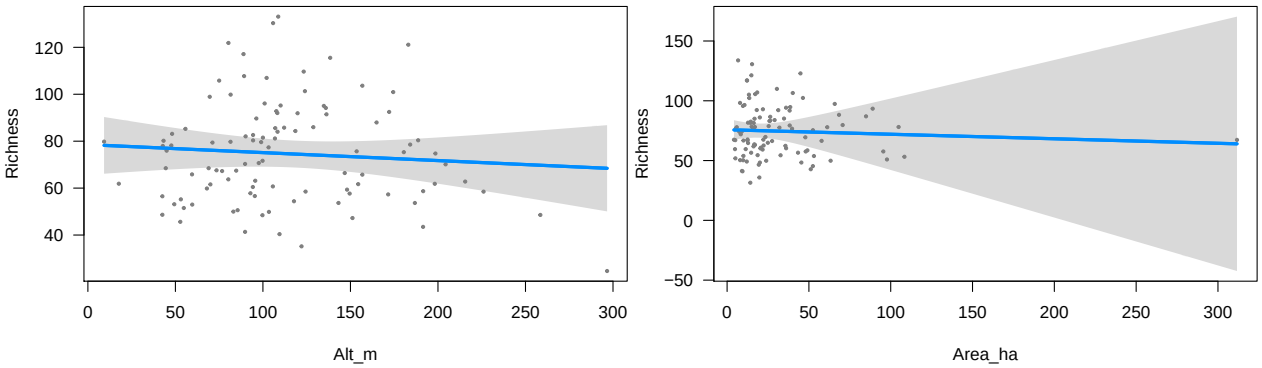
Physical Variables

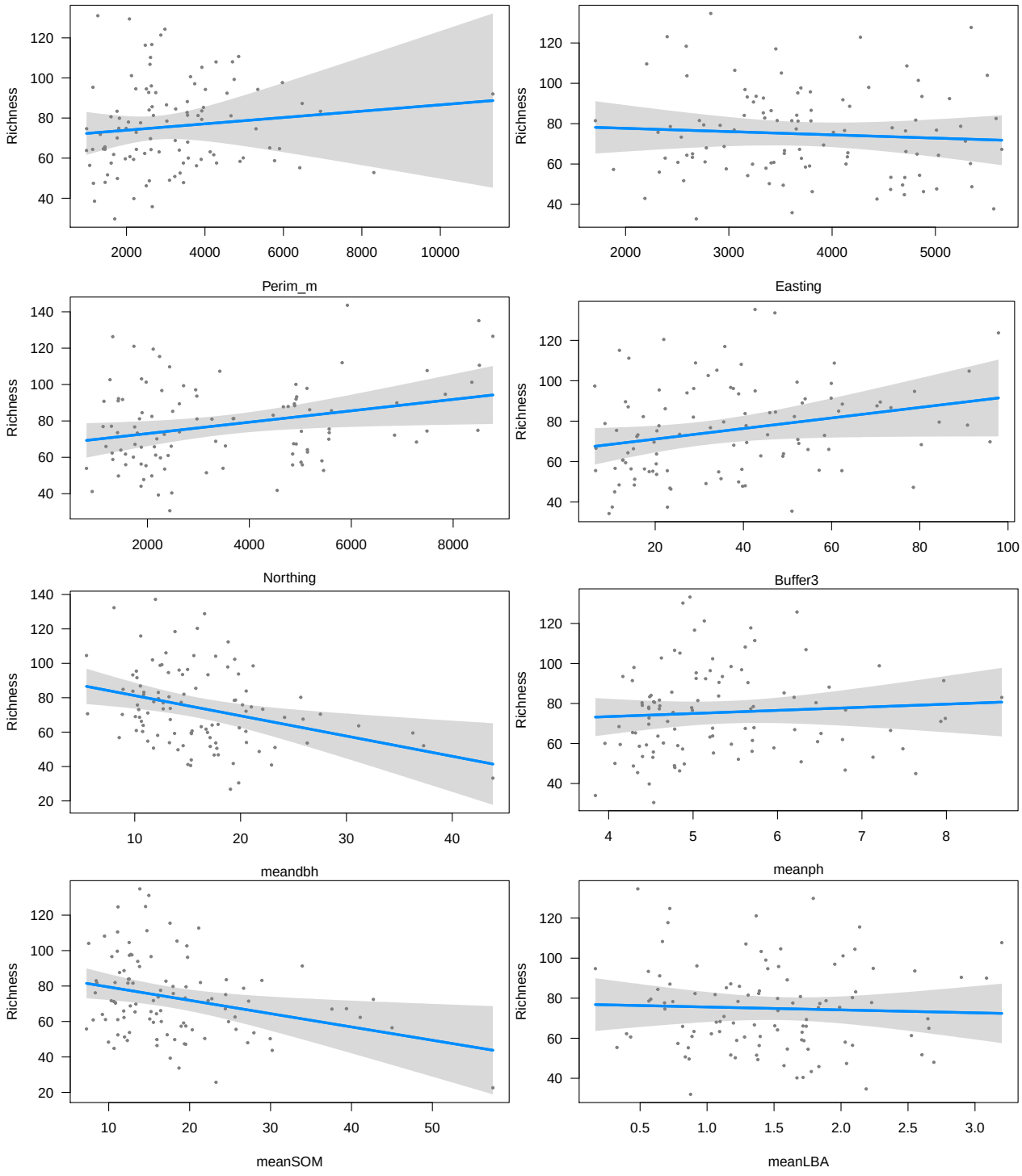
```
##  [1] "Alt_m"      "Area_ha"    "Perim_m"    "Easting"    "Northing"
##  [6] "Buffer3"    "meandbh"    "meanph"     "meanSOM"    "meanLBA"
## [11] "area_ratio"
```

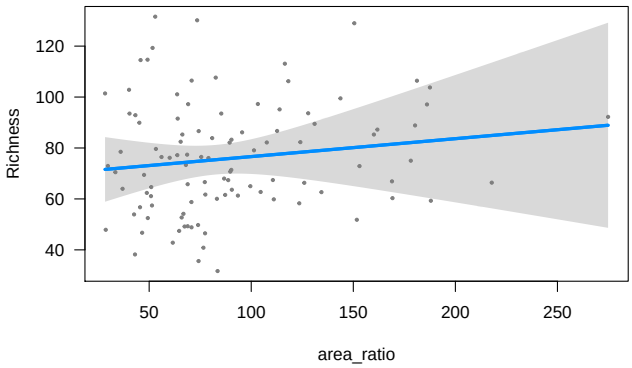
Heterogeneity variables

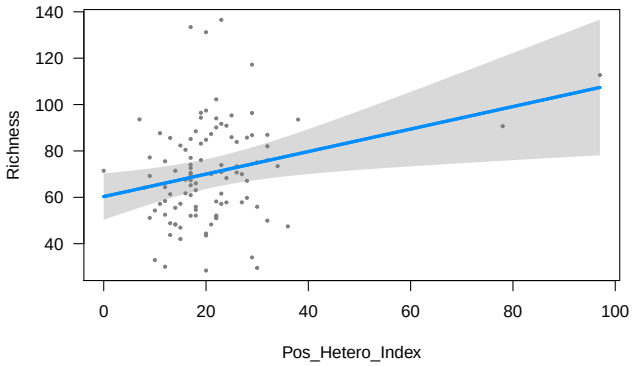
```
## [1] "Pos_Hetero_Index" "no_NVC"           "sd_pH"
## [4] "sd_SOM"           "no_MSG"           "sd_LBA"
## [7] "sd_meandbh"       "sd_treedensity"   "no_trees"
```

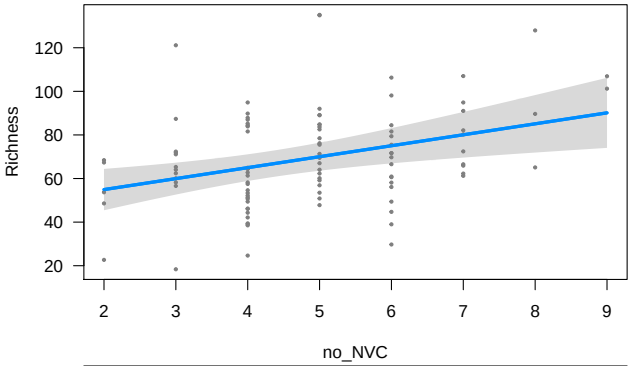
Effect of each variable on richness using multiple linear regression, simple additive model.

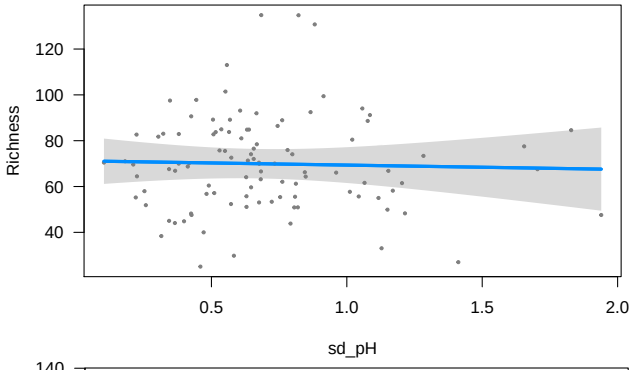


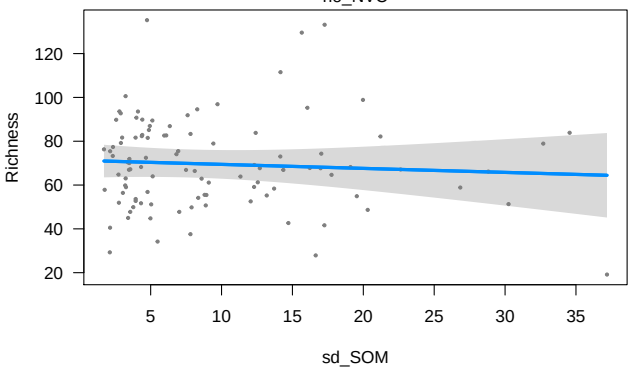

Richness
Perim_m
Easting
Northing
Buffer3
meandbh
meanph
meanSOM
meanLBA

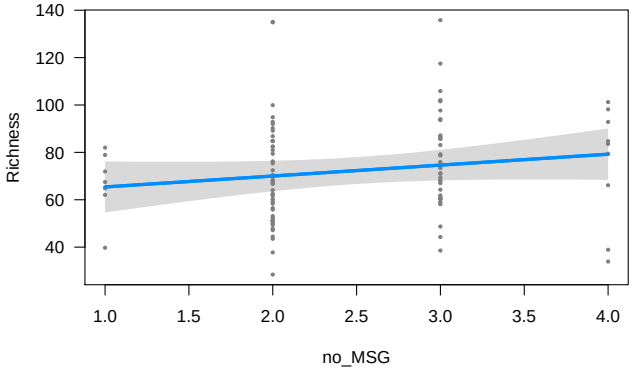

Richness
area_ratio

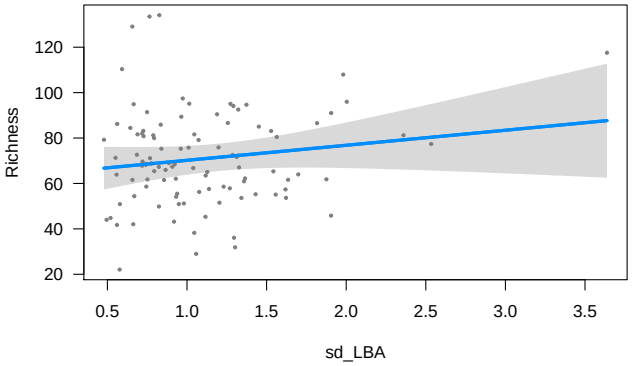

Richness
Pos_Hetero_Index

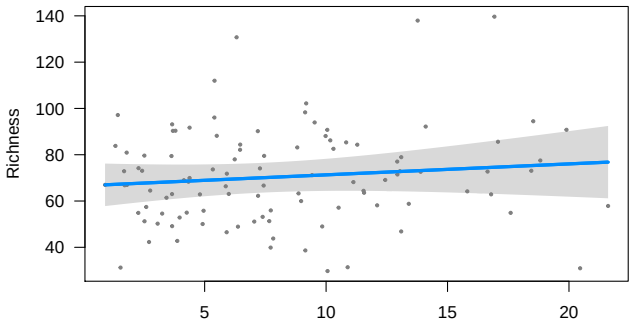

Richness
no_NVC

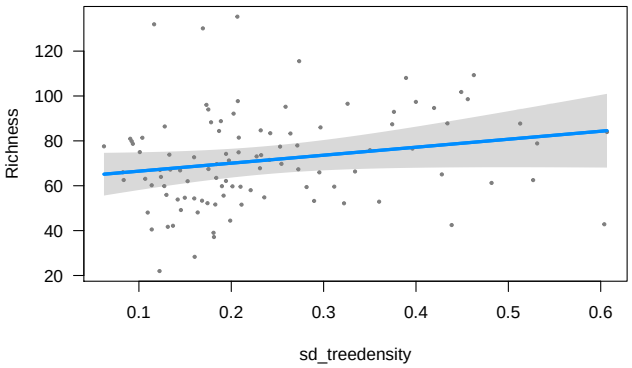

Richness
sd_pH

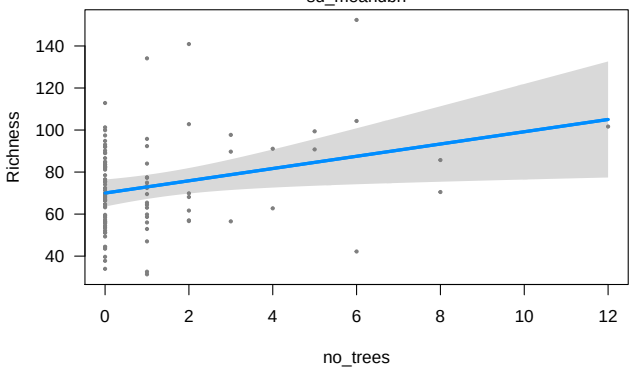

Richness
sd_SOM


Richness
no_MSG

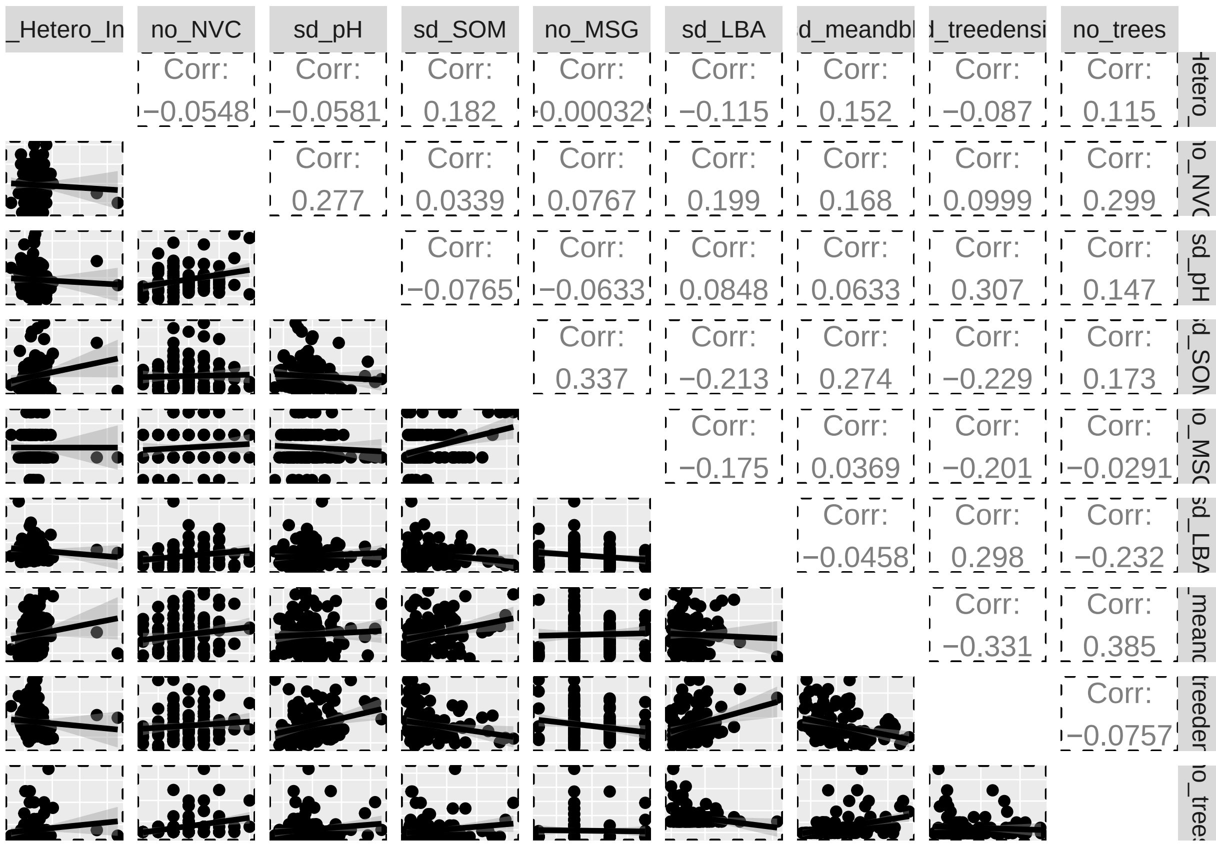

Richness
sd_LBA


Richness
140
120
100
80
60
40
5
10
15
20
sd_meandbh


Richness
120
100
80
60
40
20
0.1
0.2
0.3
0.4
0.5
0.6
sd_treedensity


Richness
140
120
100
80
60
40
0
2
4
6
8
10
12
no_trees

Colinearity of variables using pair plots



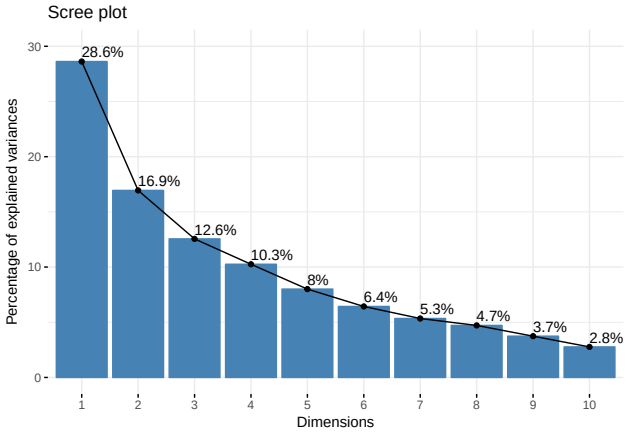
_Hetero_In
no_NVC
sd_pH
sd_SOM
no_MSG
sd_LBA
d_meandb
_treedensi
no_trees
Corr: −0.0548
Corr: −0.0581
Corr: 0.182
Corr: 0.00032
Corr: −0.115
Corr: 0.152
Corr: −0.087
Corr: 0.115
Corr: 0.277
Corr: 0.0339
Corr: 0.0767
Corr: 0.199
Corr: 0.168
Corr: 0.0999
Corr: 0.299
Corr: −0.0765
Corr: −0.0633
Corr: 0.0848
Corr: 0.0633
Corr: 0.307
Corr: 0.147
Corr: 0.337
Corr: −0.213
Corr: 0.274
Corr: −0.229
Corr: 0.173
Corr: −0.175
Corr: 0.0369
Corr: −0.201
Corr: −0.0291
Corr: −0.0458
Corr: 0.298
Corr: −0.232
Corr: −0.331
Corr: 0.385
Corr: −0.0757

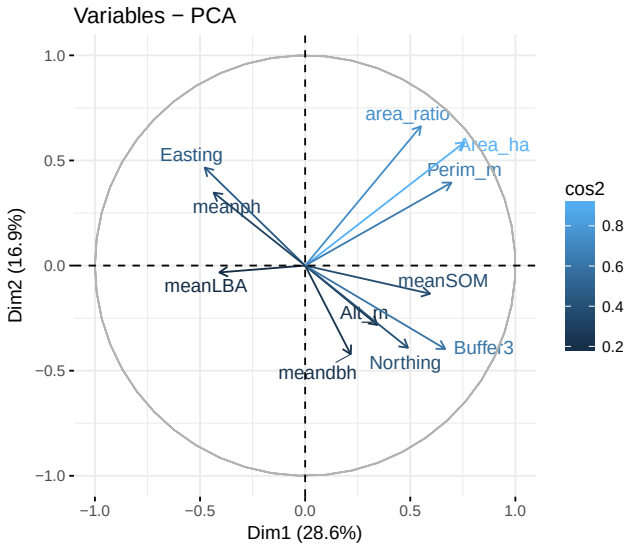
PCA

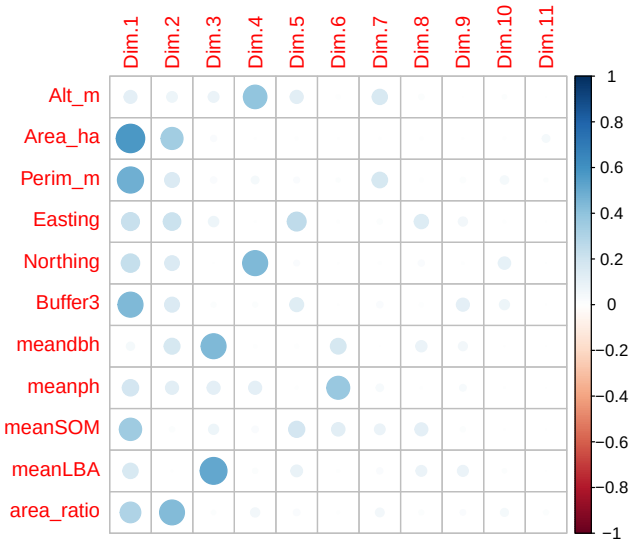
Physical variables

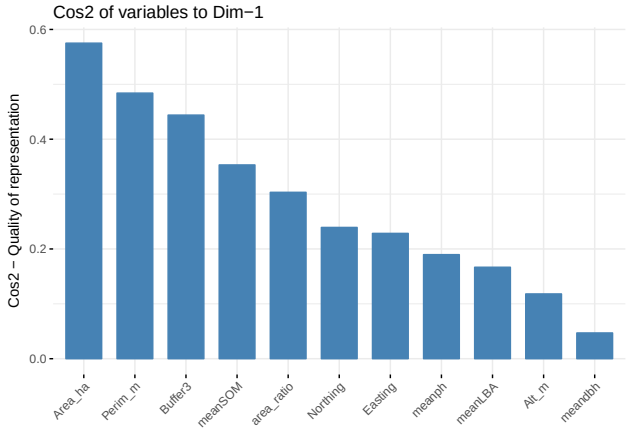
```
##        eigenvalue variance.percent cumulative.variance.percent
## Dim.1  3.14838141        28.621649                    28.62165
## Dim.2  1.86382228        16.943839                    45.56549
## Dim.3  1.38082861        12.552987                    58.11848
## Dim.4  1.12804353        10.254941                    68.37342
## Dim.5  0.88048220         8.004384                    76.37780
## Dim.6  0.70750145         6.431831                    82.80963
## Dim.7  0.58740467         5.340042                    88.14967
## Dim.8  0.51850848         4.713713                    92.86339
## Dim.9  0.41134544         3.739504                    96.60289
## Dim.10 0.30448566         2.768051                    99.37094
## Dim.11 0.06919627         0.629057                   100.00000
```

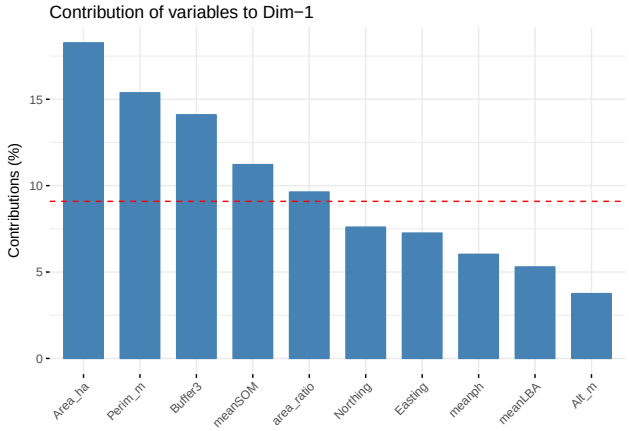
5 components explain 76% of the variance, the first 2 explain 45%

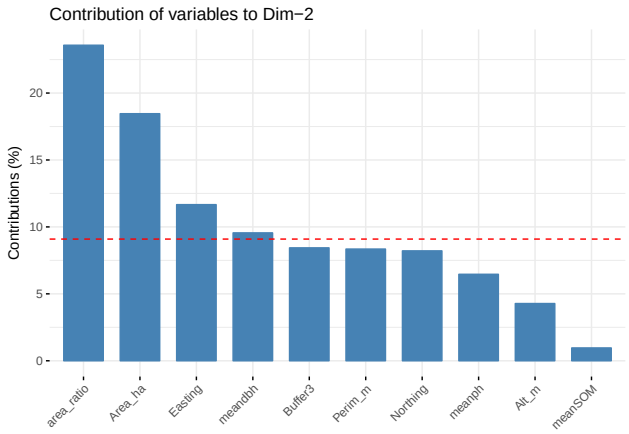





Dim.1
Dim.2
Dim.3
Dim.4
Dim.5
Dim.6
Dim.7
Dim.8
Dim.9
Dim.10
Dim.11
Alt_m
Area_ha
Perim_m
Easting
Northing
Buffer3
meandbh
meanph
meanSOM
meanLBA
area_ratio
1
0.8
0.6
0.4
0.2
0
−0.2
−0.4
−0.6
−0.8
−1

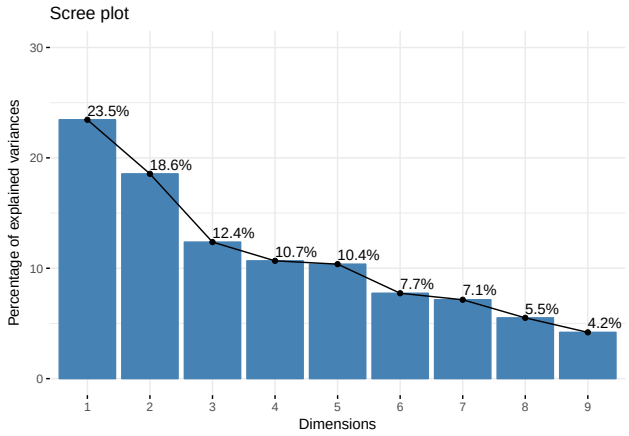

Cos2 of variables to Dim−1
Cos2 − Quality of representation
0.6
0.4
0.2
0.0
Area_ha
Perim_m
Buffer3
meanSOM
area_ratio
Northing
Easting
meanph
meanLBA
Alt_m
meandbh

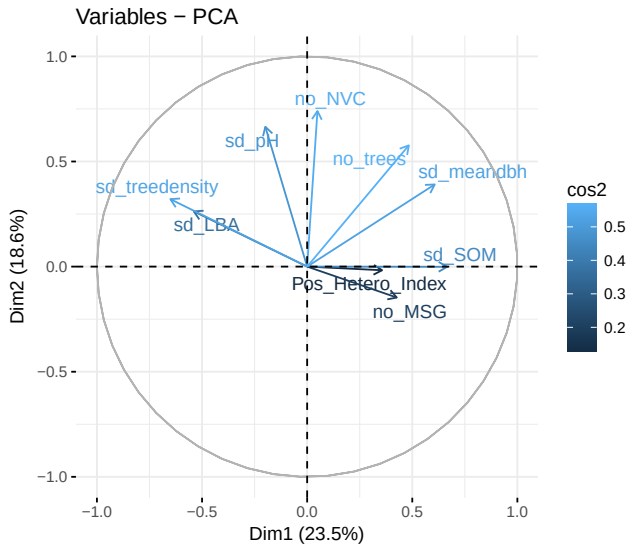

Contribution of variables to Dim−1
Contributions (%)
15
10
5
0
Area_ha
Perim_m
Buffer3
meanSOM
area_ratio
Northing
Easting
meanph
meanLBA
Alt_m

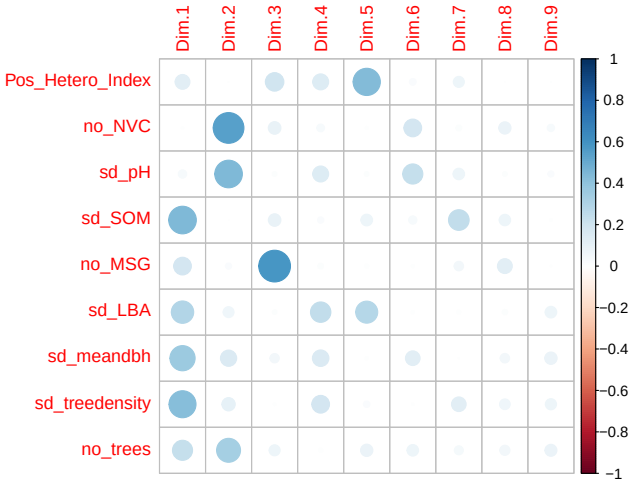


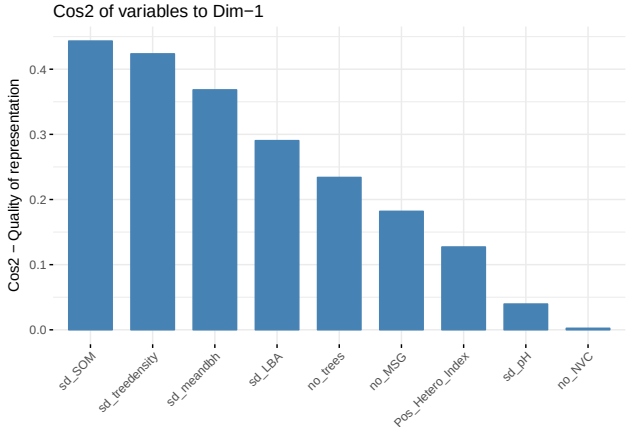

```
##       eigenvalue variance.percent cumulative.variance.percent
## Dim.1  2.1106576        23.451751                    23.45175
## Dim.2  1.6697278        18.552532                    42.00428
## Dim.3  1.1133576        12.370640                    54.37492
## Dim.4  0.9603886        10.670984                    65.04591
## Dim.5  0.9334267        10.371407                    75.41731
## Dim.6  0.6961738         7.735265                    83.15258
## Dim.7  0.6431857         7.146508                    90.29909
## Dim.8  0.4956866         5.507629                    95.80672
## Dim.9  0.3773956         4.193285                   100.00000
```

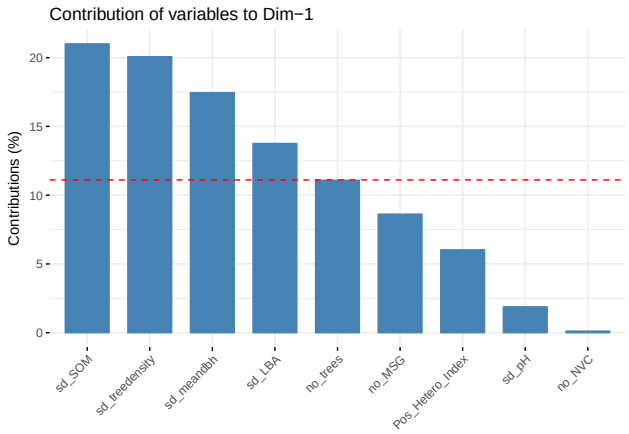
75% of variance is expressed by the first 5 components, 41% by the first 2

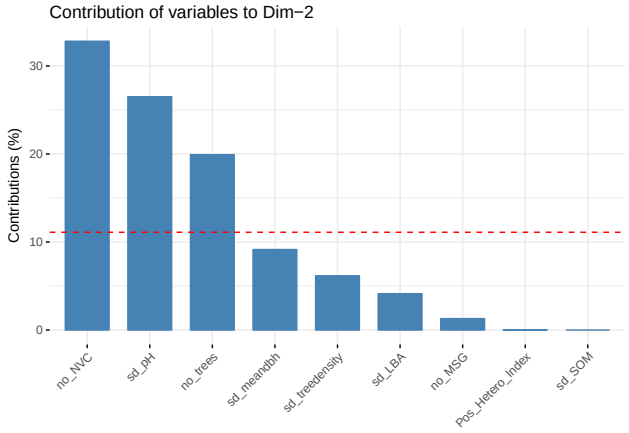



Variables − PCA
no_NVC
sd_pH
no_trees
sd_meandbh
sd_treedensity
sd_LBA
sd_SOM
Pos_Hetero_Index
no_MSG
cos2
0.5
0.4
0.3
0.2
1.0
0.5
0.0
−0.5
−1.0
Dim2 (18.6%)
Dim1 (23.5%)


Dim.1
Dim.2
Dim.3
Dim.4
Dim.5
Dim.6
Dim.7
Dim.8
Dim.9
Pos_Hetero_Index
no_NVC
sd_pH
sd_SOM
no_MSG
sd_LBA
sd_meandbh
sd_treedensity
no_trees
1
0.8
0.6
0.4
0.2
0
−0.2
−0.4
−0.6
−0.8
−1


Cos2 of variables to Dim−1
Cos2 − Quality of representation
0.4
0.3
0.2
0.1
0.0
sd_SOM
sd_treedensity
sd_meandbh
sd_LBA
no_trees
no_MSG
Pos_Hetero_Index
sd_pH
no_NVC





Regression

The plots show that the following variables have a positive correlation with species richness : Northing, Buffer, mean SOM, area to perimeter ratio ,Number of positive heterogeneity indices,number of NVC codes, number of sites with no trees, standard deviation of the tree density.

Mean dbh has a negative corrleation.

The correlation with Positive heterogenity index may be heavily influenced by the two outliers, Site 23 with 97 and Site 53 with 78),

Pairs Plots

Abiotic variables: No correlation is greather than 0.52 between any variables, accept area and perimeter and area ratio - as you'd expect. Buffer and Northing have correlation coefficient of 0.52 and Easting and mean SOM of -0.417."

Heterogenity variables: The largest correlation coefficient is 0.385, between standard deviation of mean dbh and number of sites with no trees, not unexpected. Number of major soil groups and standard devation of SOM have correlation coefficient of 0.337 and number of NVC codes and number of plots with no trees has 0.299."

PCA

Abiotic : Area, perimeter, buffer contribute to the first component, area ratio and area to the second, mean dbh and mean LBA to the third, northing and altitude to the fourth. The corr plot shows area, perimeter, buffer and mean soil organic matter increse together, as domean dbh and mean LBA, as well as Northing and altitude.

The strong contribution of area and perimeter may be due to the outlier - one wood had a area of over 300ha,(Site 74) the others were all below 150ha.

The inclusion of area, perimeter and area/perimeter ratio may seems strange, but they may have different effects on richness. A large area may increase richness due to increased species pool. A large area/perimeter ratio may increase richness due to the increased possibility of species migration, or conversely have a negative impact because the wood may then not contain a core of undisturbed woodland with poorly dispersing woodland species, which are then lost. Buffer may be a good substitute for all the above variables because it also represents a species pool, a wood with a large perimeter and a large area will equally have a large buffer.

Heterogeneity: The standard deviation of soil organic matter, tree density, LBA and mean dbh contribute the first principle component. Number of NVC codes, sd of soil pH and number of sites with no trees to the second and number of MSG to the third.

The corr plot shows sd SOM increases with sd of tree density and sd of mean dbh, and that number of NVC codes increase with sd of soil pH.

Discussion

Abiotic vars. Area, perimeter and area ratio show correlations above 0.7, as you would expect. Only area has strong relationship with richness, but this may be due to the outlier. Of the three variables perimeter is not required as it is the area, or the geometry of the wood, expressed in the area/preimeter ratio that might affect richness. Therefore the outlier will be removed and the perimeter will be removed.

The other correlations are all below 0.5 and there isnt a clear reason why any should be dropped, e.g. buffer and northing are not ecologically related, so it doesnt make sense to drop one at this stage.

PCA shows mean_dbh and mean_lba increase together, and they are obviously related, so one of these coud be dropped. Its not obvious why they are not more strongly correlated??

DROP?? perimeter, meanLBA, outlier

Heterovars. Number of sites with no trees and sd mean dbh show correlation of 0.385 - since they are comutationally related both may be unnecessary. PCA shows that sd_treedensity increases with sd_mean dbh - again related variables, it might be worth dropping both since mean dbh which has a strong influence on richness, and is being used. Since number of sites with no trees has the strongest influence on richness, keep this variable and drop se_meandbh, sd_treedensity, sd_meandbh.

Number of MSG and sd MSG groups, likewise

PS = Check correlation between no_trees and mean dbh you would expect the to be related, but they werent considered her because of the way the variables ahve been split into two groups.

DROP?? sd meandbh, sd_treedensity, sd_MSG, outliers