ExaminingCovariatesVI

PetraGuy

27 March 2018

From previous analysis it became obvious (maybe should have been in the first place) that most of the variables I had "created" in order to express heterogeneity were already contained within the existing variates. Eg, the standard deviation of pH and number of major soil groups. In addition, I wasnt sure what standardising a standard deviation in the PCA would mean? And - there were so many variables the analysis was becoming harder rather than simpler - none of the additional variables added obvious correlation with richness. So I decided to strip the variables back to those that were originally measured and recorded.

Here I will go through these and remove firstly, those that are not correlated with species richess amd/or are correlated with each other - eg. we dont want all of area, permimeter and area_ratio, but which is most important for richness? It has been suggested that the geometry of the wood can be related to richness - eg, an undistrbed core could contain poorly dispersing woodland species, or a long narrow wood would more easily allow recolonisation. So arearatio might be the most useful variable.

There are also a few outliers that may affect the analysis, one site(74) has area > 300ha, the others area less that 150ha. Including this site gives a stronger correlation with richness and area than when it is removed. Two sites have positive heterogeneity indices above 78, the rest are below 40. Again, the correlation with richness is reduced if they are removed.

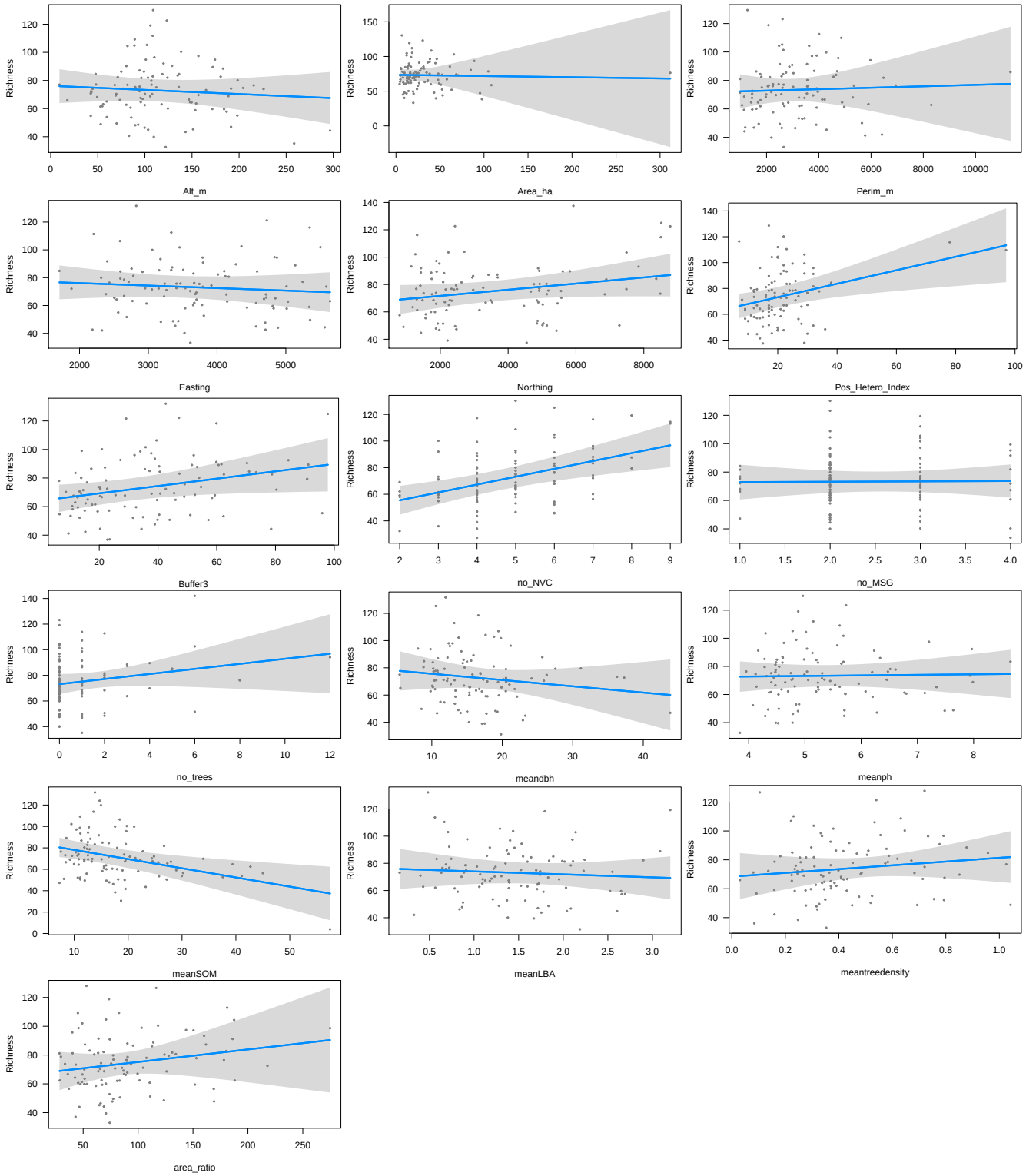
In previous analysis I split the variables in two - physical variables (eg mean dbh) and heterogeneoty variables. The problem with this was that correlations across the two groups, in the presence of the other variables could not be seen.

```
##  [1] "X"                "Site"             "Alt_m"
##  [4] "Area_ha"          "Perim_m"          "Easting"
##  [7] "Northing"         "No.of.Des"        "Any.Anc"
## [10] "Pos_Hetero_Index" "Buffer1"          "Buffer2"
## [13] "Buffer3"          "Richness"         "no_NVC"
## [16] "sd_pH"            "sd_SOM"           "no_MSG"
## [19] "sd_LBA"           "sd_meandbh"       "sd_treedensity"
## [22] "no_trees"         "sd_R"             "mean_R"
## [25] "sd_N"             "mean_N"           "sd_W"
## [28] "mean_W"           "sd_L"             "mean_L"
## [31] "meandbh"          "meanph"           "meanSOM"
## [34] "meanLBA"          "meantreedensity"  "area_ratio"
```

These were all the variables I created because I was to express the heterogenity of the woodland. But none showed a correaltion with richness.

```
##  [1] "Alt_m"            "Area_ha"          "Perim_m"
##  [4] "Easting"          "Northing"         "Pos_Hetero_Index"
##  [7] "Buffer3"          "Richness"         "no_NVC"
## [10] "no_MSG"           "no_trees"         "meandbh"
## [13] "meanph"           "meanSOM"          "meanLBA"
## [16] "meantreedensity"  "area_ratio"
```

These are the measured variables - I have calculated mean - which is the mean over all the plots - plots with no trees have a dbh of 0 and therefore this variable also reflects the heterogeneity of the woodland in relation to its cover. This means that number of sites with no trees (no_trees), meanLBA and tree density are probably not all required.



The multiple linear regression using all variables shows the following variables are correlated with richness : Buffer,Northing, no_NVC, no_trees, meandbh, meanSOM, meantreedensity, arearatio. (It is a bit surprising that meanLBA is not included in the correlations with tree variables, perhaps there are too many tree variables?)

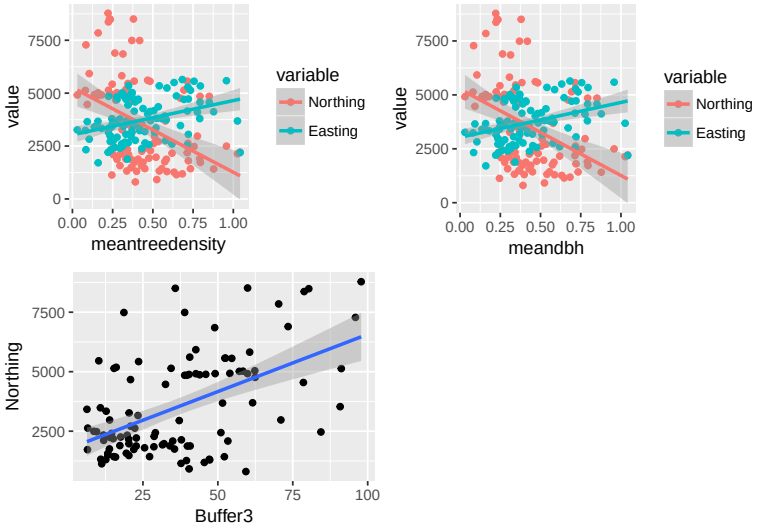
Area,perimeter and area ratio, are correlated, as you'd expect. Also Buffer and Northing (0.52) Do Northern

woods have larger buffers because there's more room?

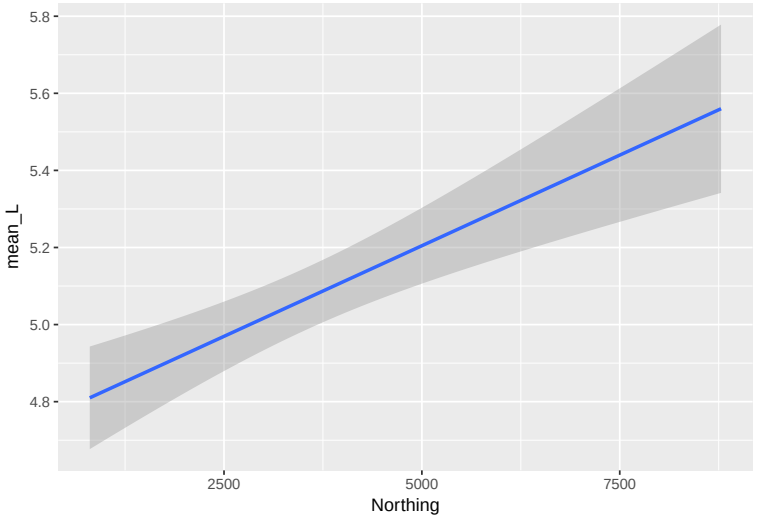
Also mean tree density, meanLBA, meanDBH no_trees, (0.306, -0.442, -0.417,) also as you'd expect, since all are also computationally similar. But meantreedensity is also correlated with buffer, Northing and Easting (-0.38, -0.416, 0.336)

This implies that Eastern woods are denser while northern woods are more open, and woods with a larger buffer are less dense - yes, because they are northern.Buffer is correlated with northing (0.52), then what we are seeing is the expression of the less dense northern woods.

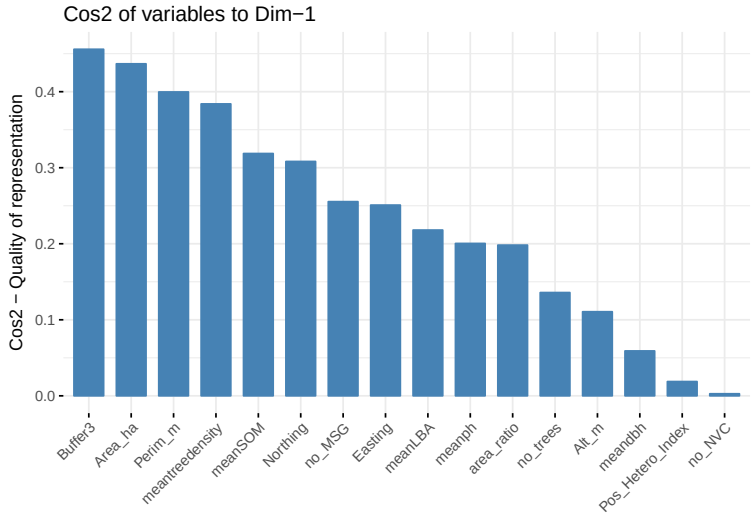
That is, northern woods both have a larger buffer, **and** are less dense. This may be a problem when looking at the correaltion with richness - is the increase in richness related to the density, the buffer or the latitude - there are ecological reasons why all could be true. Less dense woods allow the introduction of more light loving species. A larger buffer provides a seed bank, allows recolonisation and offers potential protection from nitrogen enrichment and the further north you go, the less the species richness, (not sure how much this counts in UK - is the change in latitude enough? Is it more likely to be due to soil types etc which would swamp any latitude effect).



If the less dense northern woods also had more plants with higher L Ellenberg values, this might suggest that the increase in richness was related to the tree variables.



The increase in the mean_L with Northing supports the idea that it is the reduced density (or mean dbh) of the woodland that is adding to the species richness, and not the buffer (or the Latitude?)

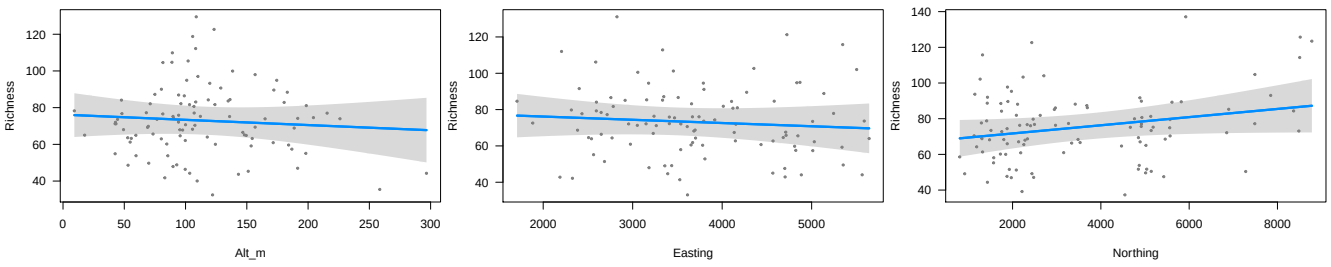


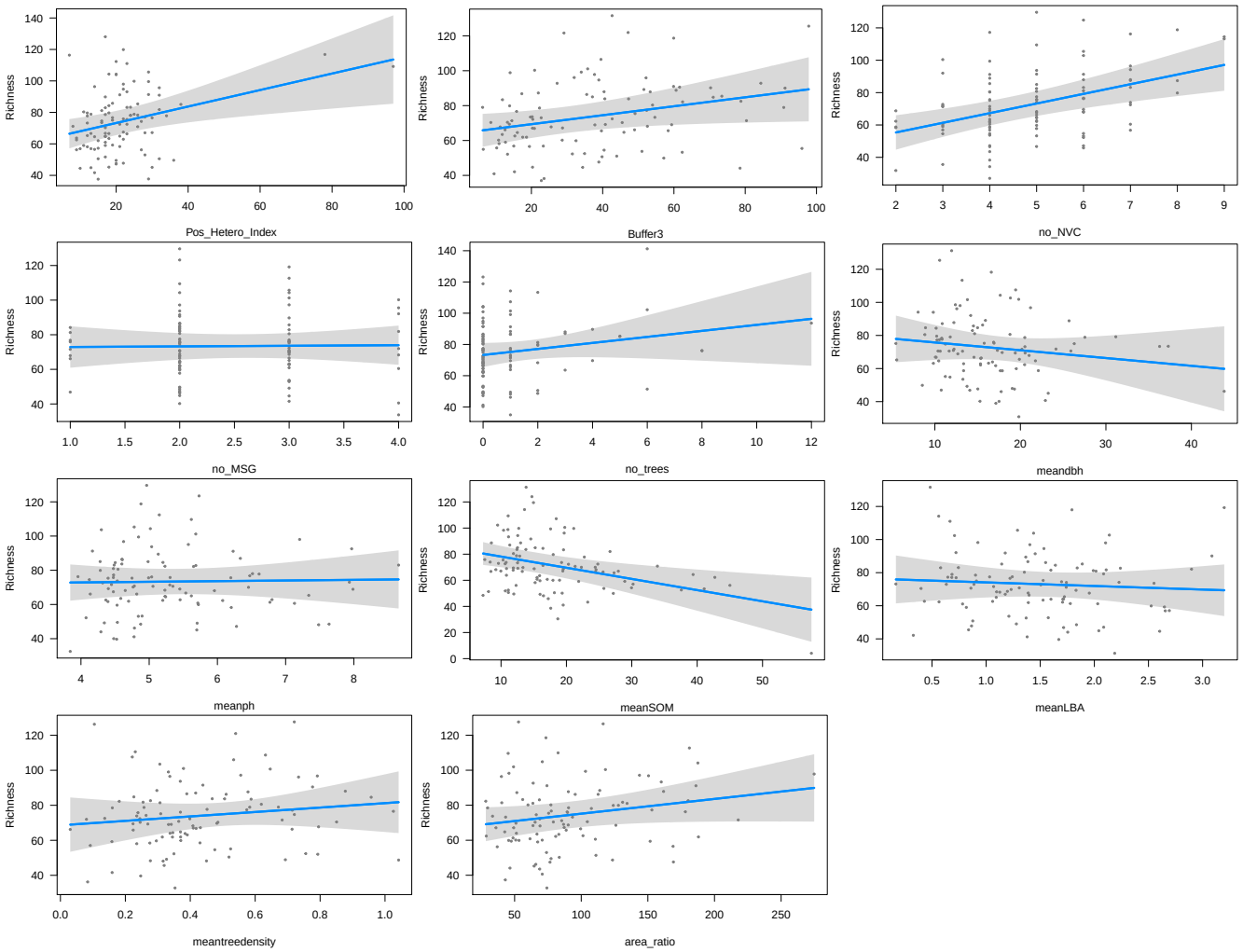
Area and Perimeter contribute to the variance, could this be due mainly to the outlier? The mean tree density contributes more than the meandbh - No idea what that means, especially since mean dbh has a greater gradient with richness than meantreedensity.

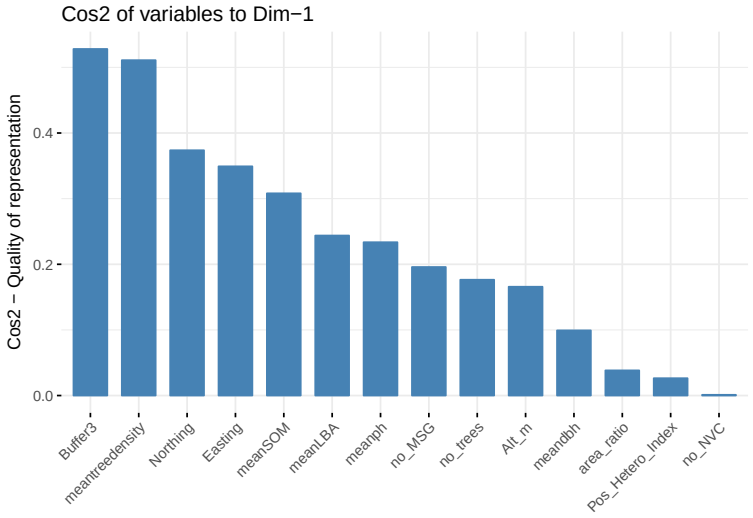
No_NVC also has no contribution to variance, but has strong correlation with richness.

```
##                      p   R2
## (Intercept)       0.18 0.45
## Alt_m             0.52    -
## Area_ha           0.92    -
## Perim_m           0.83    -
## Easting           0.51    -
## Northing           0.1    -
## Pos_Hetero_Index 0.01    -
## Buffer3           0.04    -
## no_NVC               0    -
## no_MSG            0.94    -
## no_trees          0.18    -
## meandbh           0.35    -
## meanph            0.87    -
## meanSOM              0    -
## meanLBA           0.62    -
## meantreedensity    0.4    -
## area_ratio        0.36    -
```

Removing area and perimeter







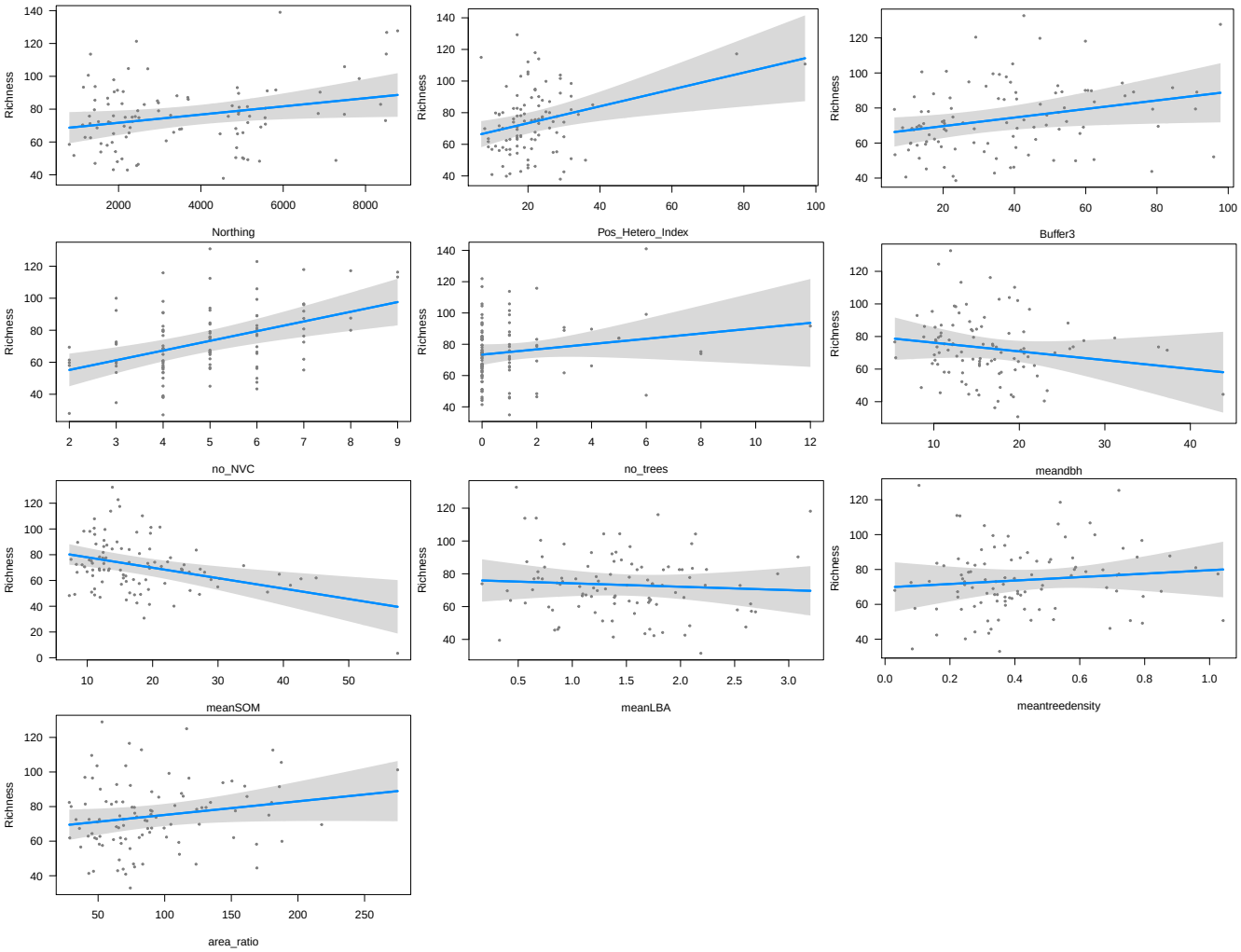

```
##                    p    r
## (Intercept)     0.14 0.44
## Alt_m           0.52    -
## Easting          0.5    -
## Northing        0.08    -
```

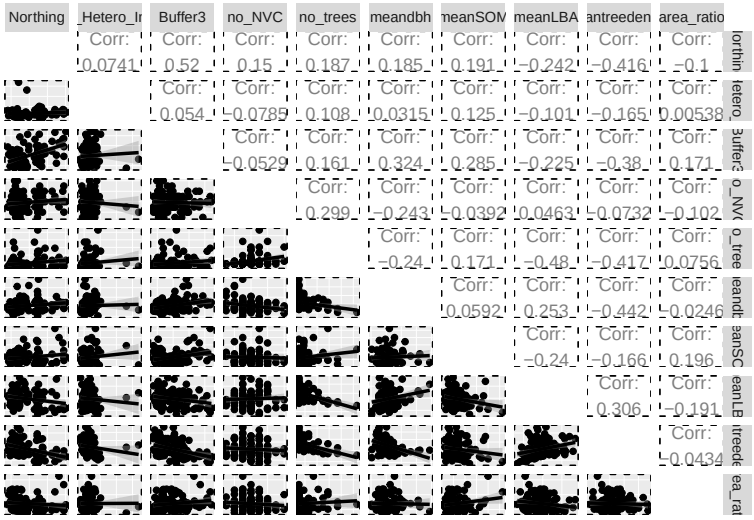
```
## Pos_Hetero_Index    0    -
## Buffer3          0.04    -
## no_NVC              0    -
## no_MSG           0.91    -
## no_trees         0.18    -
## meandbh          0.33    -
## meanph           0.87    -
## meanSOM             0    -
## meanLBA          0.62    -
## meantreedensity   0.4    -
## area_ratio       0.08    -
```

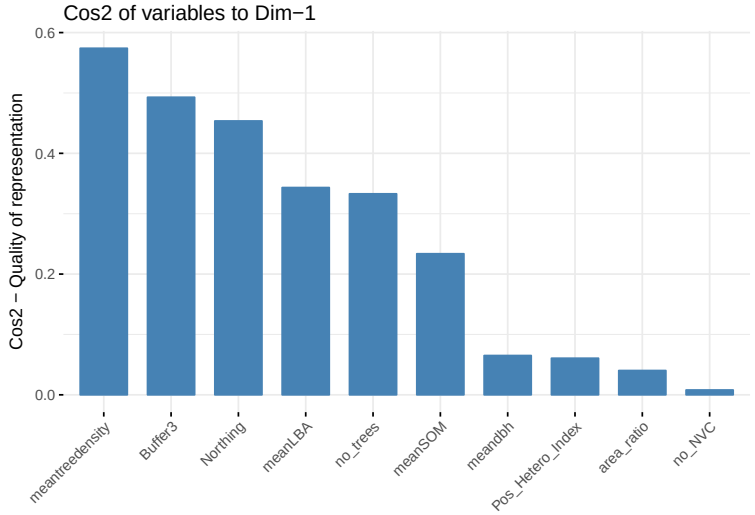
Removing area and perimeter doesnt effect the correlation with richness (accept to make area_ratio a bit more signifiant).

Droppping Easting, number MSG,mean pH and altitude

None of these variables contribute to the richness and none are significant. We already saw that any contribution asting might make is via tree density.



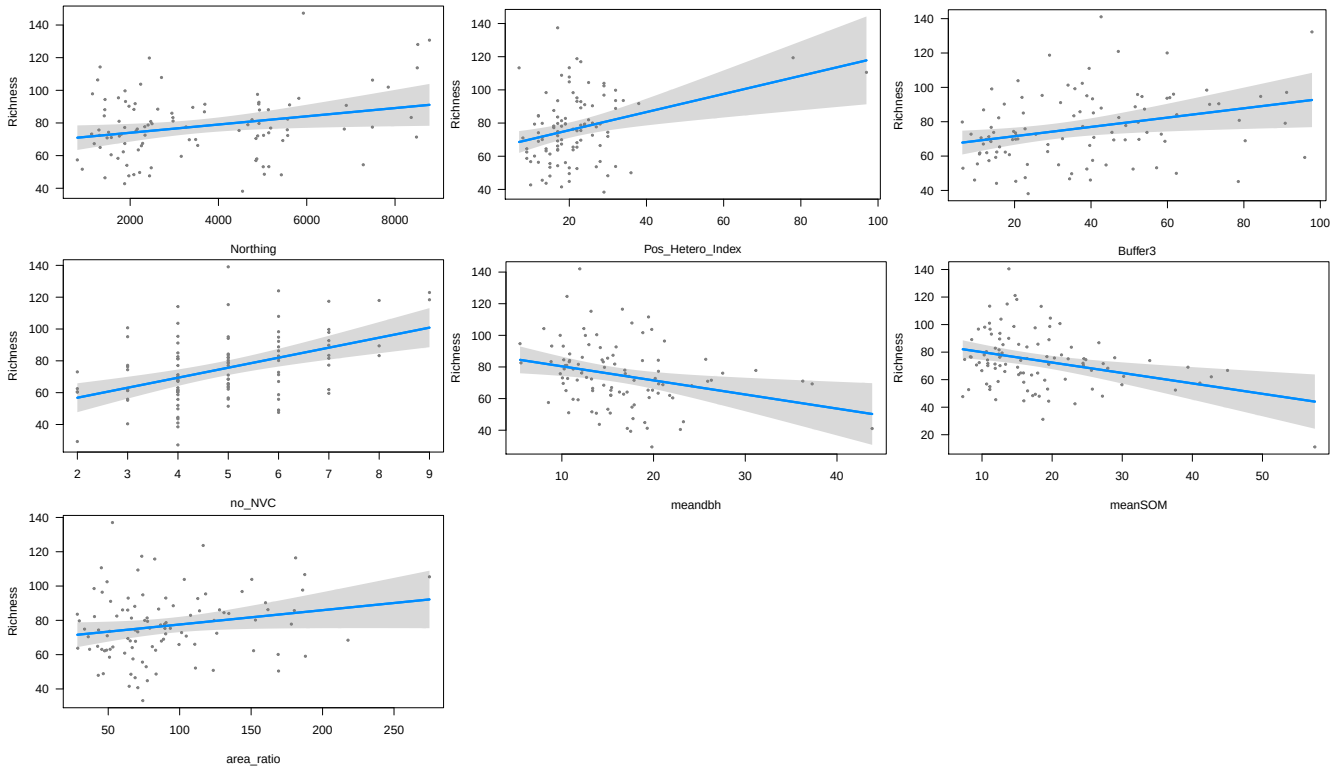


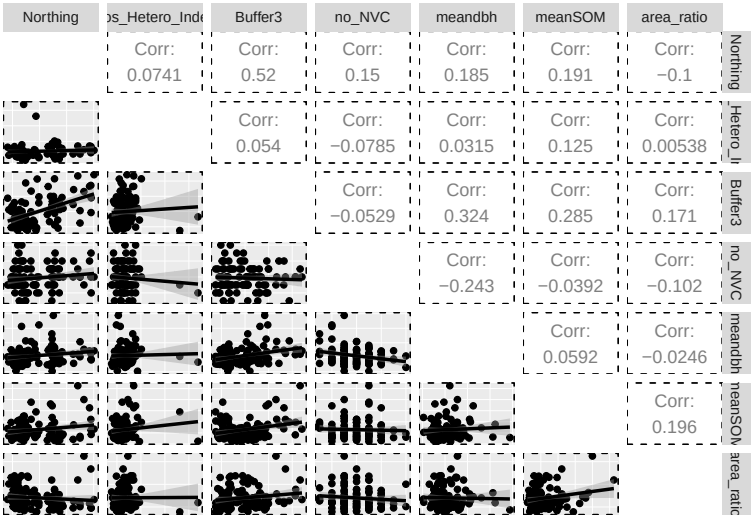


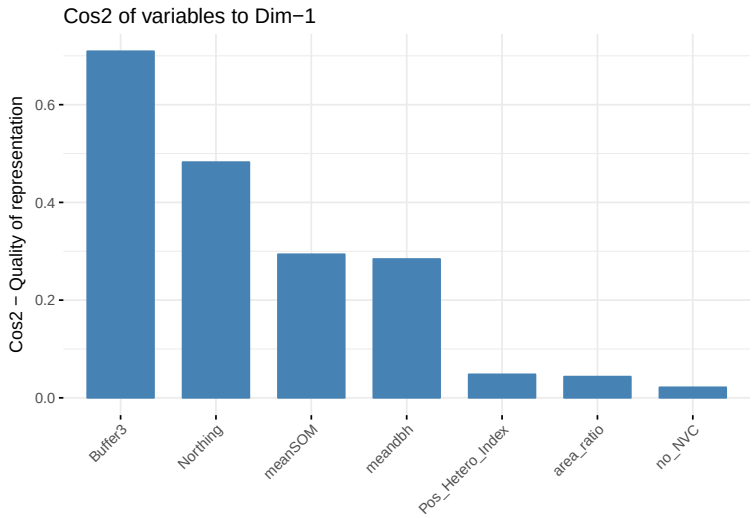

```
##                       p    r
## (Intercept)        0.07 0.44
## Northing           0.04    -
## Pos_Hetero_Index      0    -
## Buffer3            0.04    -
## no_NVC                0    -
## no_trees           0.21    -
## meandbh            0.25    -
## meanSOM               0    -
## meanLBA            0.61    -
## meantreedensity    0.48    -
## area_ratio         0.09    -
```

Drop meanLBA, mean tree density, no_trees

Since we dont need all the tree variables, and meandbh has the strongest correlation with richness.


Richness
Northing
Pos_Hetero_Index
Buffer3
meandbh
meanSOM
no_NVC
area_ratio

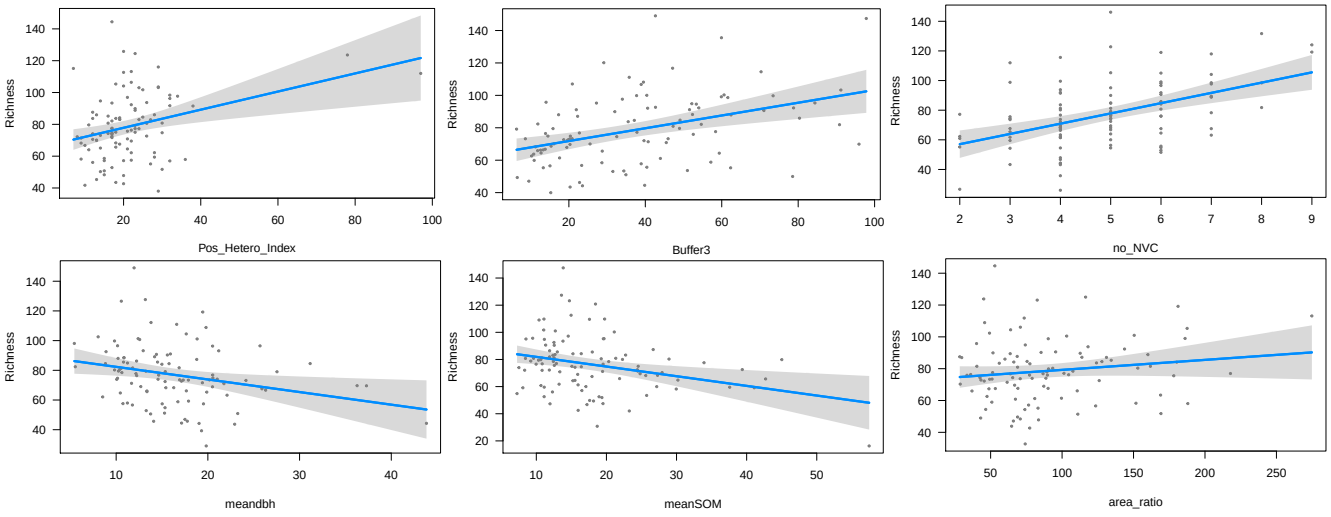

Northing
Pos_Hetero_Index
Buffer3
no_NVC
meandbh
meanSOM
area_ratio
Corr: 0.0741
Corr: 0.52
Corr: 0.15
Corr: 0.185
Corr: 0.191
Corr: −0.1
Corr: 0.054
Corr: −0.0785
Corr: 0.0315
Corr: 0.125
Corr: 0.00538
Corr: −0.0529
Corr: 0.324
Corr: 0.285
Corr: 0.171
Corr: −0.243
Corr: −0.0392
Corr: −0.102
Corr: 0.0592
Corr: −0.0246
Corr: 0.196

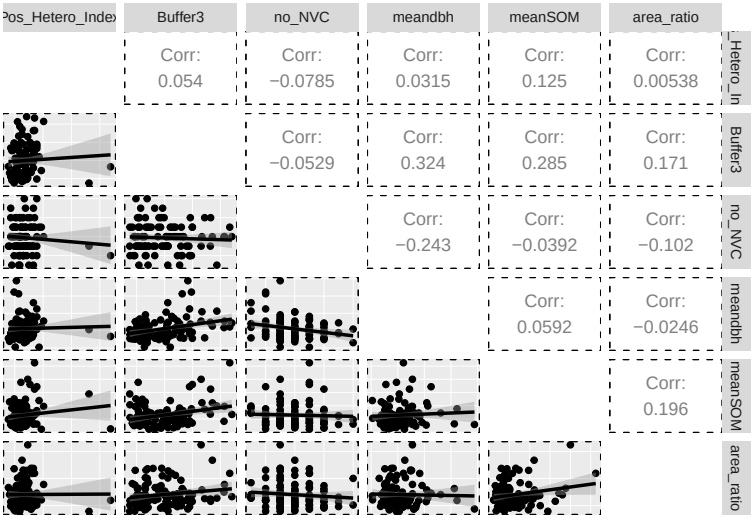


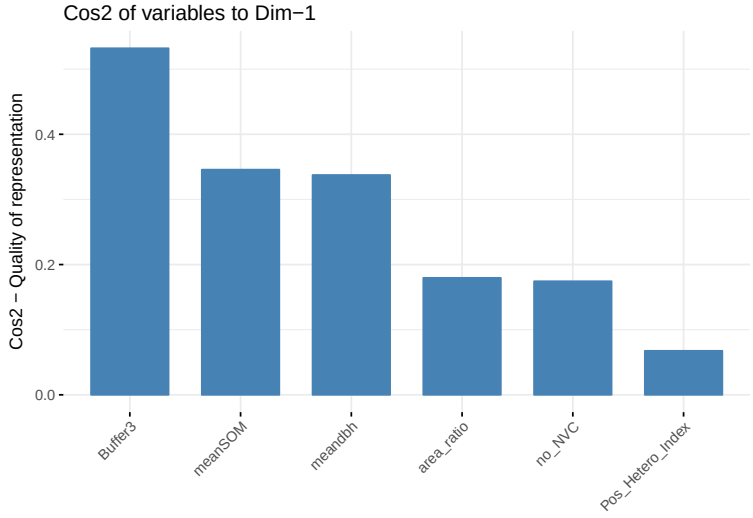

```
##                      p    r
## (Intercept)          0 0.42
## Northing          0.03    -
## Pos_Hetero_Index     0    -
## Buffer3           0.02    -
## no_NVC               0    -
## meandbh           0.01    -
## meanSOM              0    -
## area_ratio        0.07    -
```

If the correlation with richness is about tree density - for which I have retained meandbh, then buffer and Northing are spurious. Lets look at fits without first Northing, Then Buffer, then both.

Remove Northing keep Buffer

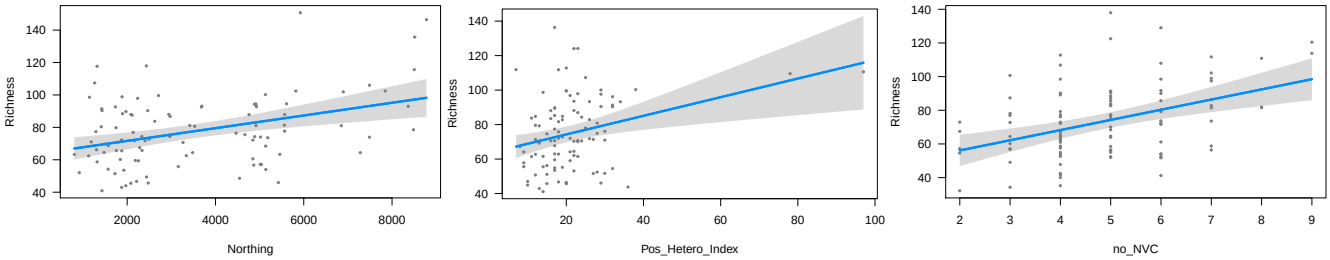


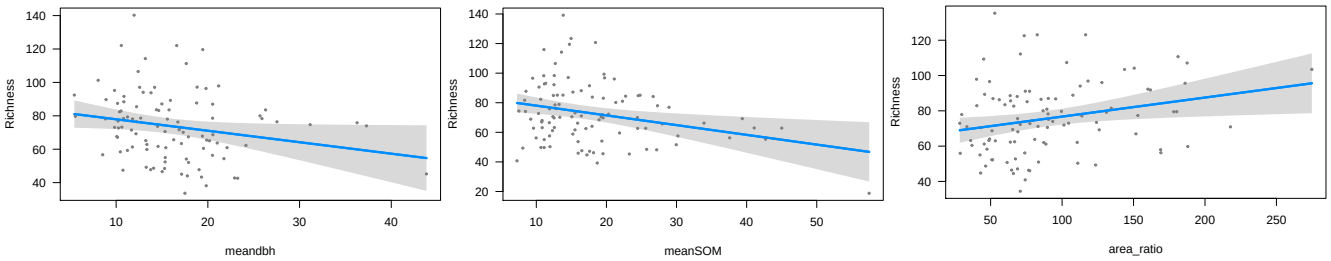


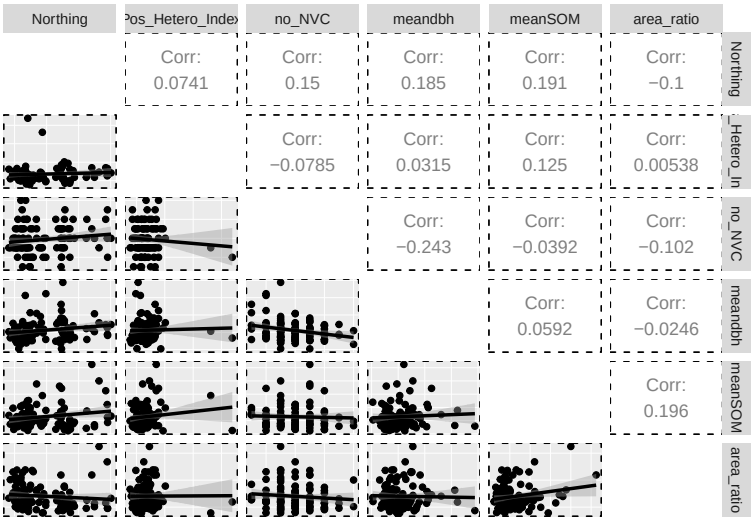



```
##                      p   r
## (Intercept)          0 0.4
## Pos_Hetero_Index     0   -
## Buffer3              0   -
## no_NVC               0   -
## meandbh           0.02   -
## meanSOM              0   -
## area_ratio        0.16   -
```

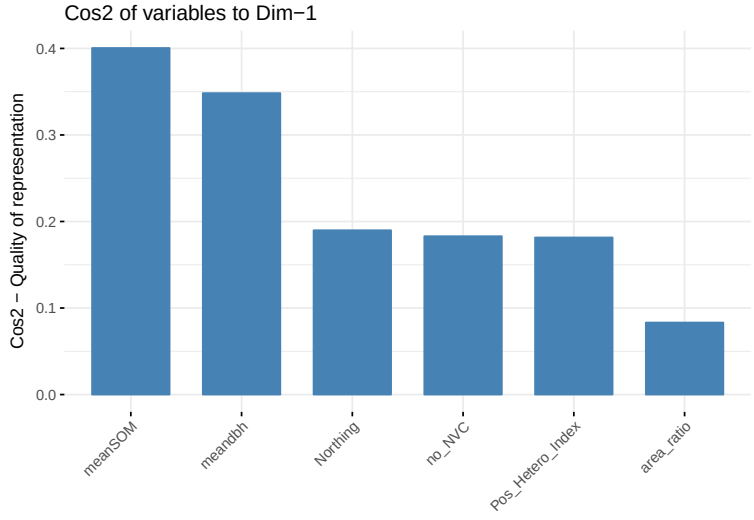
Keep Buffer remove Northing.







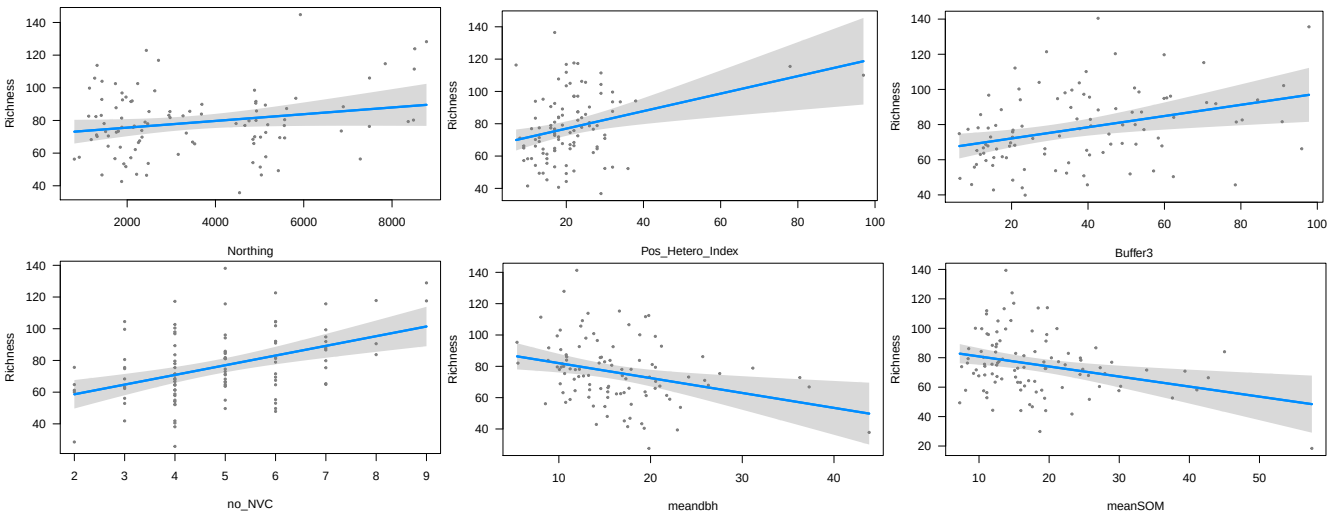

```
##       eigenvalue variance.percent cumulative.variance.percent
## Dim.1  1.3865234        23.108724                    23.10872
## Dim.2  1.2347994        20.579990                    43.68871
## Dim.3  1.1362941        18.938236                    62.62695
## Dim.4  0.9605594        16.009323                    78.63627
## Dim.5  0.6974706        11.624511                    90.26078
## Dim.6  0.5843530         9.739216                   100.00000
```

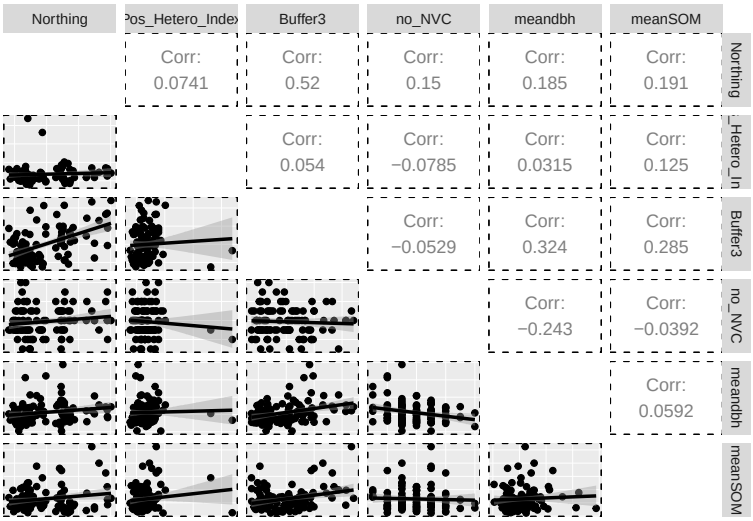


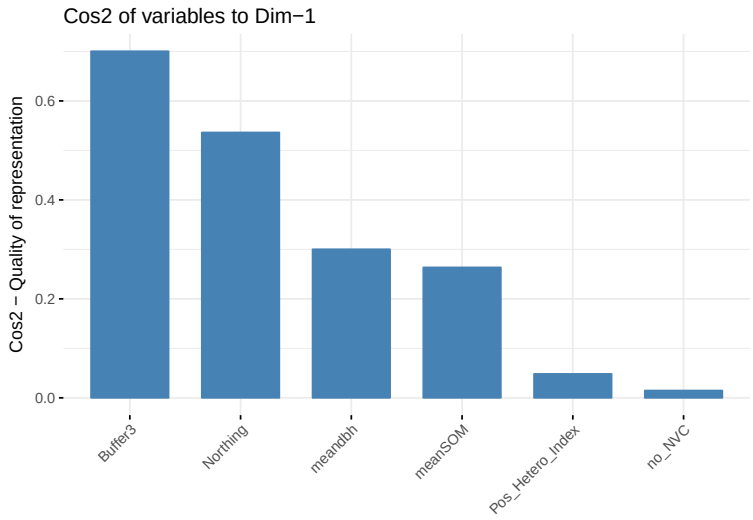

```
##                    p    r
## (Intercept)        0 0.39
## Northing           0    -
## Pos_Hetero_Index   0    -
## no_NVC             0    -
```

```
## meandbh          0.05    -
## meanSOM          0.01    -
## area_ratio       0.02    -
```

Keep Northing and Buffer remove area ratio

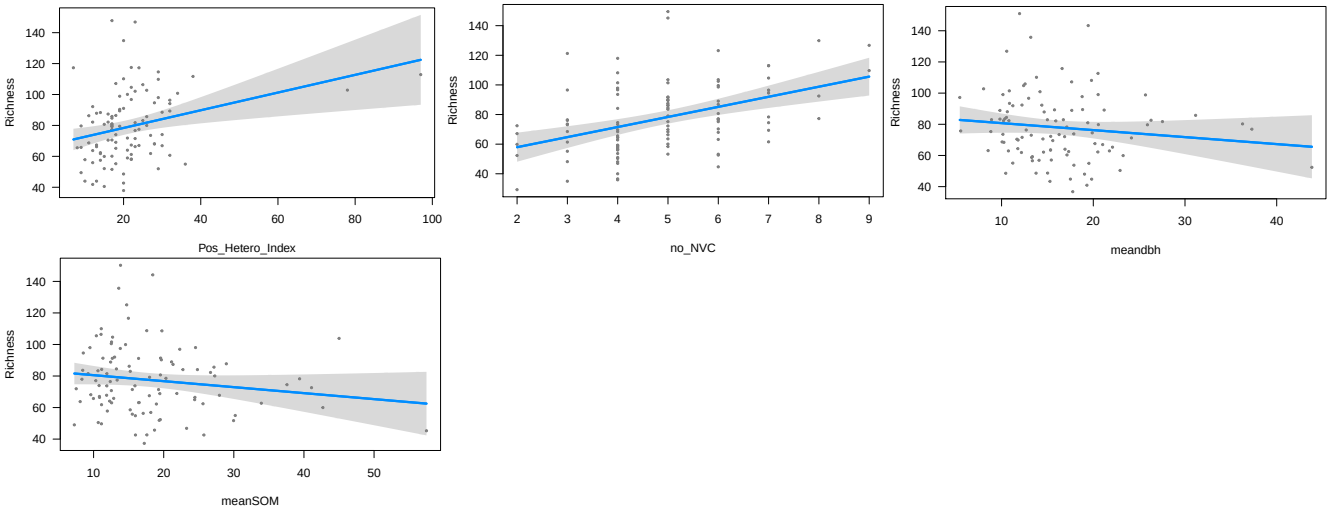


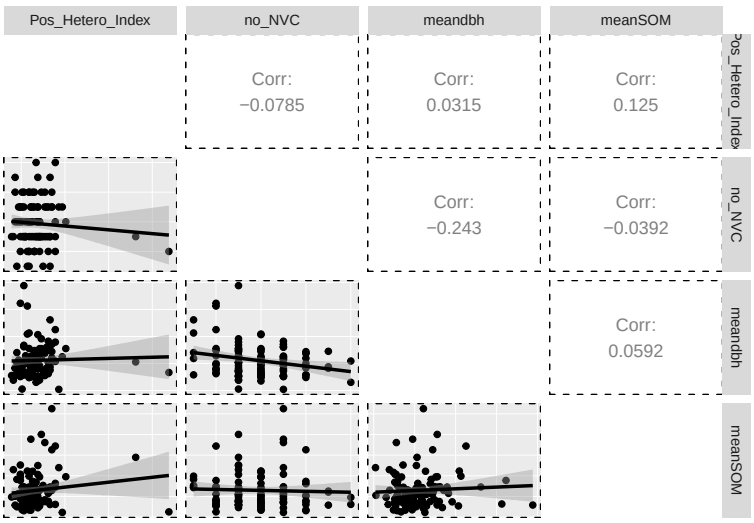


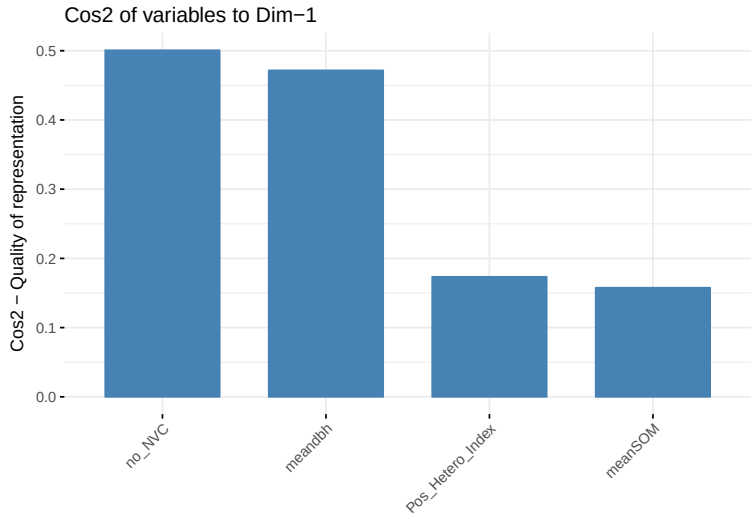



```
##                      p   r
## (Intercept)          0 0.4
## Northing          0.08   -
## Pos_Hetero_Index     0   -
## Buffer3           0.01   -
## no_NVC               0   -
## meandbh           0.01   -
## meanSOM           0.01   -
```

Remove area ratio, Northing and Buffer

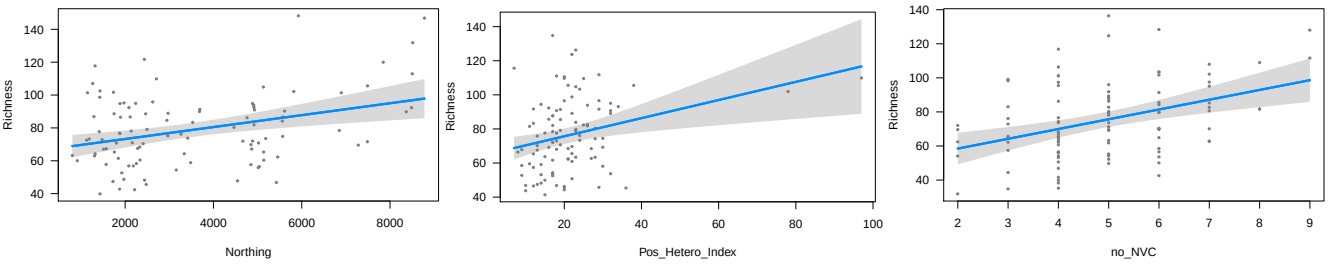


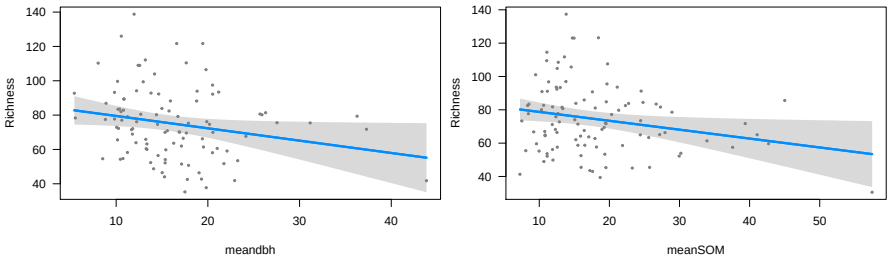


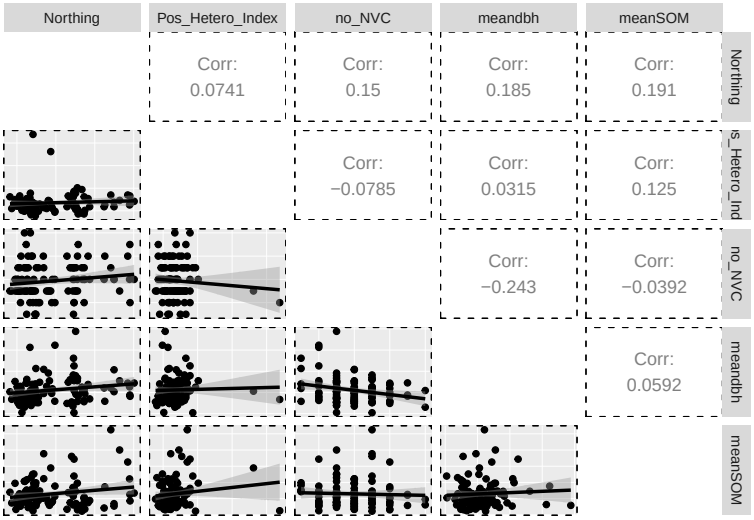


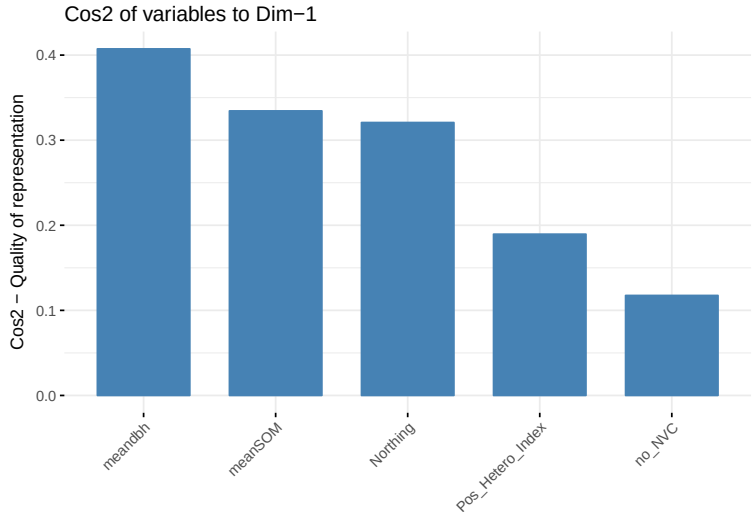

```
##                        p    r
## (Intercept)            0 0.27
## Pos_Hetero_Index       0    -
## no_NVC                 0    -
## meandbh             0.21    -
## meanSOM             0.13    -
```

With Buffer, without area ratio



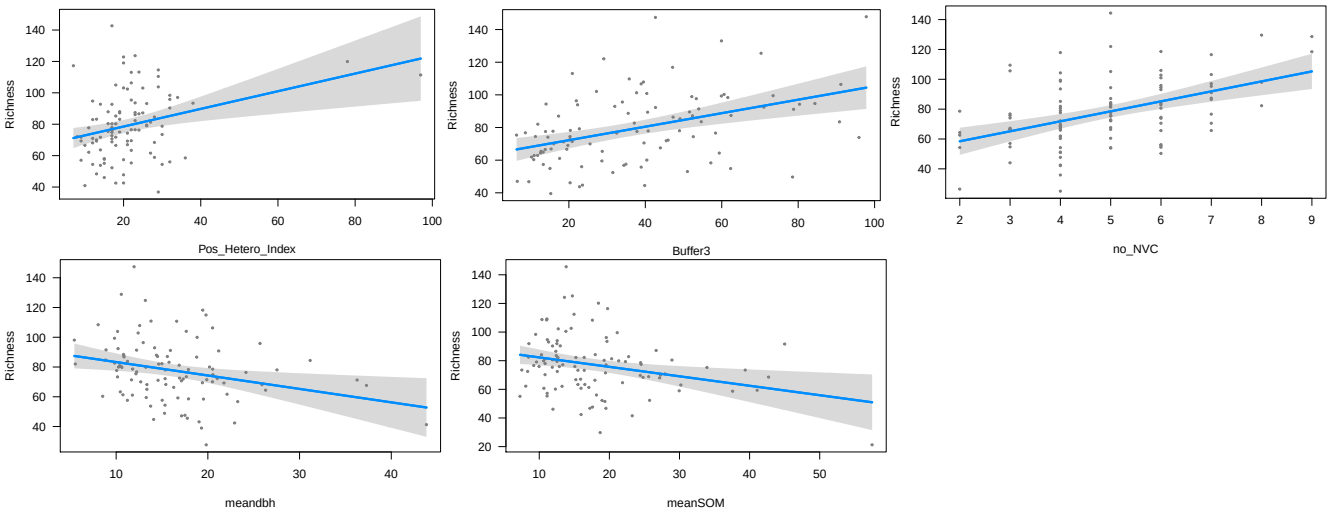


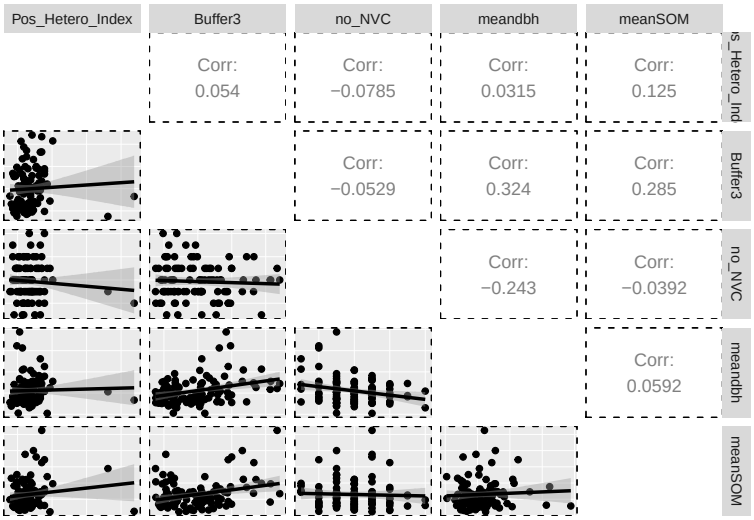


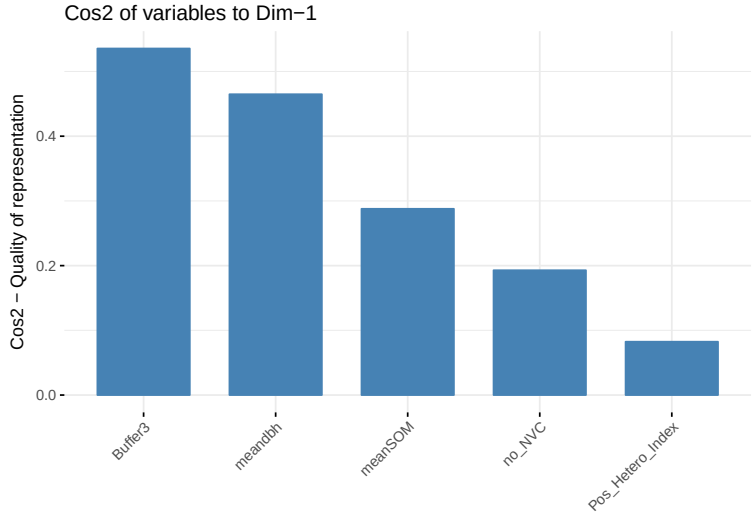



```
##                      p    r
## (Intercept)          0 0.35
## Northing             0    -
## Pos_Hetero_Index     0    -
## no_NVC               0    -
## meandbh           0.04    -
## meanSOM           0.03    -
```

With Northing without area ratio








```
##                     p    r
## (Intercept)         0 0.38
```

```
## Pos_Hetero_Index    0    -
## Buffer3             0    -
## no_NVC              0    -
## meandbh          0.01    -
## meanSOM          0.01    -
```