

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 wasn't 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. Most importantly, 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 richness and/or are correlated with each other - eg. we don't want all of area, perimeter 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 undisturbed 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 than 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 heterogeneity variables. The problem with this was that correlations across the two groups, in the presence of the other variables could not be seen.

Bearing in mind that the purpose of this analysis is to look for the factors that effect richness, and might therefore be related to z and c. It is not to create a predictive model using richness as a function of the variables. Therefore looking at R2 and AIC etc may not be best way to proceed. I tried looking at PCA, R2 and p values for various subsets, and the amount of results and graphs produced didn't add a lot of information. I have therefore decided to simplify the exploration to first looking at the multiple linear regression plots using visref because these show which variables, in the presence of the others, correlate with richness. Secondly I will look at correlations between variables using a standardised correlation matrix. These correlations will be explored to see whether the correlated variables might be confounding.

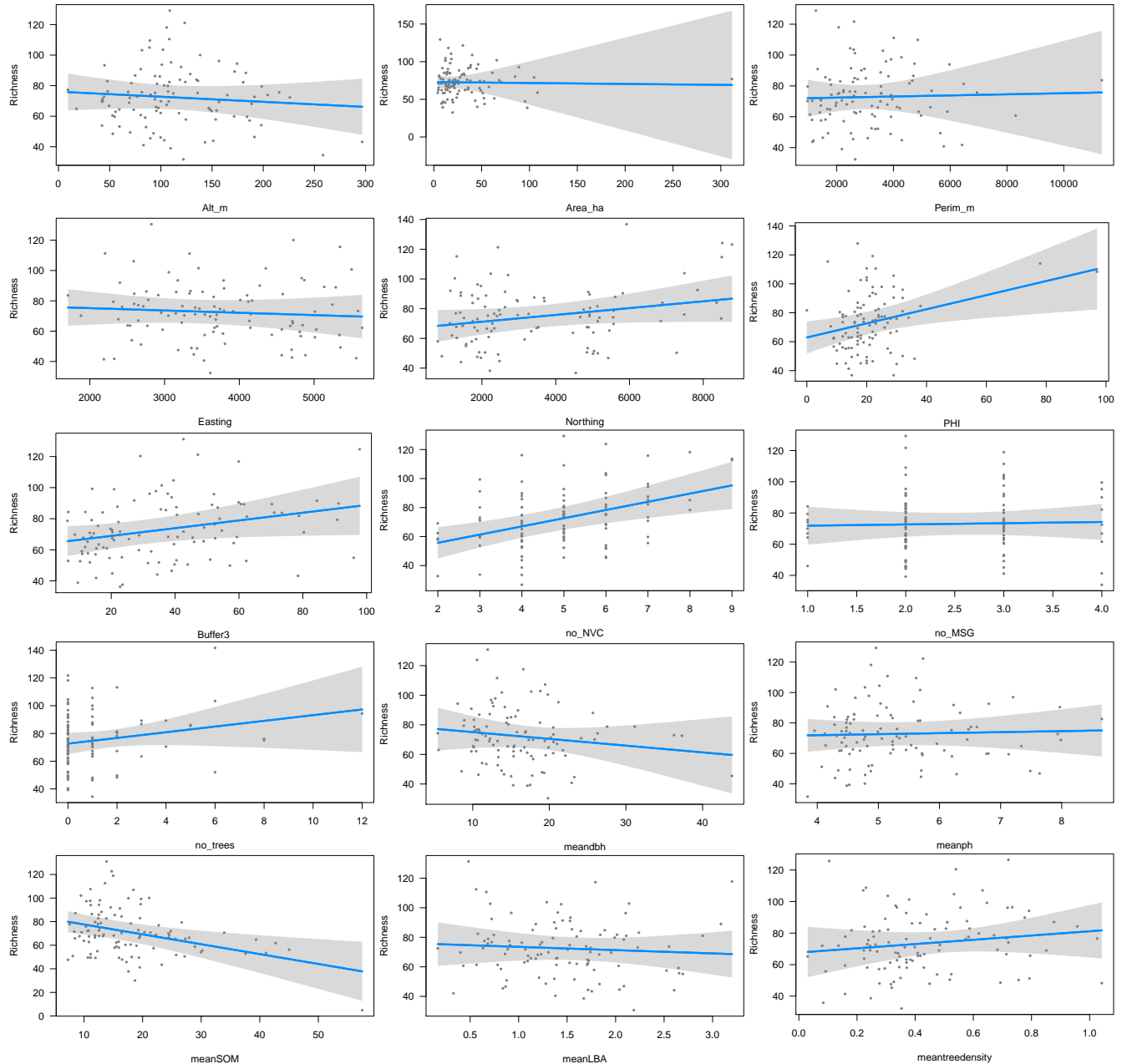
```
## [1] "X"           "Site"         "Alt_m"
## [4] "Area_ha"     "Perim_m"      "Easting"
## [7] "Northing"    "No.of.Des"    "Any.Anc"
## [10] "PHI"         "Buffer1"      "Buffer2"
## [13] "Buffer3"     "Richness"     "no_NVC"
## [16] "sd_pH"       "sd_SOM"       "no_MSG"
## [19] "sd_LBA"      "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"
```

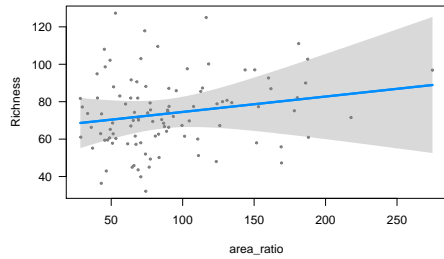
These were all the variables I created because I was to express the heterogeneity of the woodland. But none showed a correlation with richness.

```
## [1] "Alt_m"       "Area_ha"      "Perim_m"
## [4] "Easting"     "Northing"     "PHI"
## [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 not all required. Similarly, area, perimeter and area ratio are 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?)

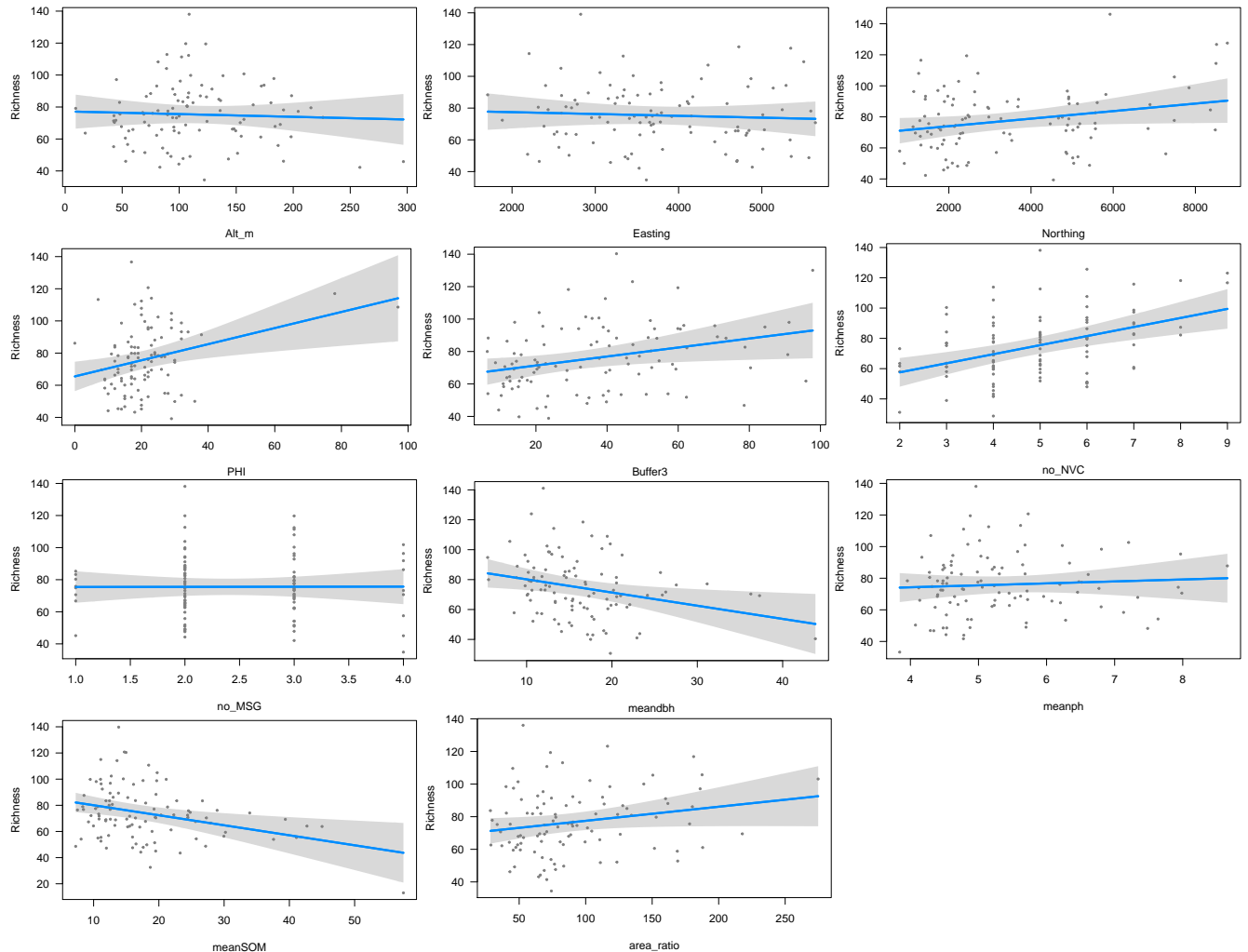
##	Alt_m	Area_ha	Perim_m	Easting	Northing	PHI	Buffer3
## Alt_m	1.00	0.03	0.04	-0.22	-0.01	-0.05	0.27
## Area_ha	0.03	1.00	0.78	-0.11	0.16	0.03	0.26
## Perim_m	0.04	0.78	1.00	-0.16	0.28	0.06	0.25
## Easting	-0.22	-0.11	-0.16	1.00	-0.27	-0.07	-0.29
## Northing	-0.01	0.16	0.28	-0.27	1.00	0.08	0.52
## PHI	-0.05	0.03	0.06	-0.07	0.08	1.00	0.08
## Buffer3	0.27	0.26	0.25	-0.29	0.52	0.08	1.00
## no_NVC	-0.06	0.00	0.14	-0.05	0.15	-0.05	-0.05
## no_MSG	0.06	0.31	0.34	-0.10	0.26	0.00	0.26
## no_trees	0.19	0.12	0.08	-0.14	0.19	0.12	0.16
## meandbh	0.08	0.03	0.03	-0.11	0.18	0.04	0.32
## meanph	-0.21	-0.14	-0.17	0.27	-0.14	-0.01	-0.36
## meanSOM	0.24	0.32	0.25	-0.42	0.19	0.13	0.29
## meanLBA	-0.24	-0.20	-0.14	0.23	-0.24	-0.10	-0.23
## meantreedensity	-0.15	-0.16	-0.20	0.37	-0.42	-0.18	-0.38
## area_ratio	0.10	0.79	0.42	0.09	-0.10	0.01	0.17
##	no_NVC	no_MSG	no_trees	meandbh	meanph	meanSOM	meanLBA
## Alt_m	-0.06	0.06	0.19	0.08	-0.21	0.24	-0.24
## Area_ha	0.00	0.31	0.12	0.03	-0.14	0.32	-0.20
## Perim_m	0.14	0.34	0.08	0.03	-0.17	0.25	-0.14
## Easting	-0.05	-0.10	-0.14	-0.11	0.27	-0.42	0.23
## Northing	0.15	0.26	0.19	0.18	-0.14	0.19	-0.24
## PHI	-0.05	0.00	0.12	0.04	-0.01	0.13	-0.10
## Buffer3	-0.05	0.26	0.16	0.32	-0.36	0.29	-0.23
## no_NVC	1.00	0.08	0.30	-0.24	0.08	-0.04	0.05
## no_MSG	0.08	1.00	-0.03	0.05	-0.37	0.25	-0.14
## no_trees	0.30	-0.03	1.00	-0.24	0.06	0.17	-0.48
## meandbh	-0.24	0.05	-0.24	1.00	-0.26	0.06	0.25
## meanph	0.08	-0.37	0.06	-0.26	1.00	-0.18	-0.04
## meanSOM	-0.04	0.25	0.17	0.06	-0.18	1.00	-0.24
## meanLBA	0.05	-0.14	-0.48	0.25	-0.04	-0.24	1.00
## meantreedensity	-0.07	-0.19	-0.42	-0.44	0.26	-0.17	0.31
## area_ratio	-0.10	0.20	0.08	-0.02	-0.10	0.20	-0.19
##	meantreedensity	area_ratio					
## Alt_m	-0.15	0.10					
## Area_ha	-0.16	0.79					
## Perim_m	-0.20	0.42					
## Easting	0.37	0.09					
## Northing	-0.42	-0.10					
## PHI	-0.18	0.01					

```
## Buffer3                -0.38      0.17
## no_NVC                 -0.07     -0.10
## no_MSG                 -0.19      0.20
## no_trees               -0.42      0.08
## meandbh                -0.44     -0.02
## meanph                  0.26     -0.10
## meanSOM                -0.17      0.20
## meanLBA                 0.31     -0.19
## meantreedensity         1.00     -0.04
## area_ratio              -0.04      1.00
```

The correlation matrix shows the expected strong correlations between area, perimeter and area ratio and meanLBA, tree density, no_trees and meandbh. Since area_ratio is potentially the more interesting, this will be retained, and since mean dbh has the strongest correlation with richness, that will also be retained.

```
## [1] "Alt_m"      "Easting"    "Northing"   "PHI"        "Buffer3"
## [6] "Richness"    "no_NVC"     "no_MSG"     "meandbh"    "meanph"
## [11] "meanSOM"     "area_ratio"
```

This is a more workable and realistic set of variables - in the sense that none are clearly correlated because they are computationally related, and they are all based on the original measured data with little manipulation.



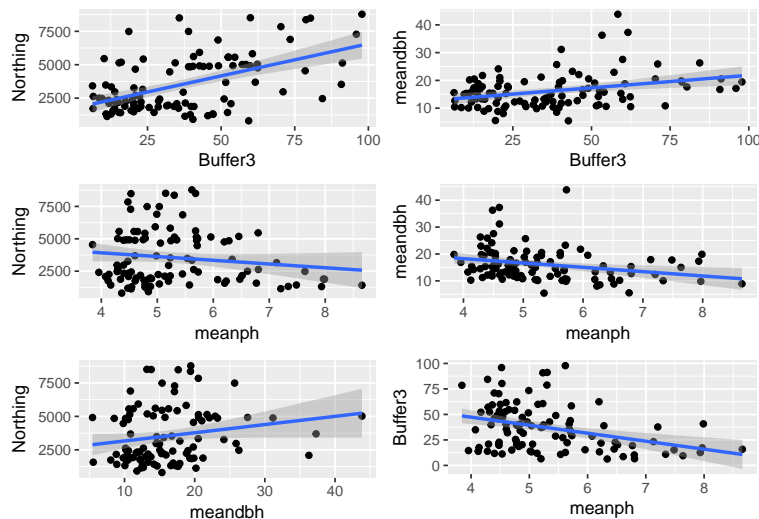
The visref plots show no correlation with richness for altitude, easting, no MSG groups or meanph

##	Alt_m	Easting	Northing	PHI	Buffer3	no_NVC	no_MSG	meandbh
## Alt_m	1.00	-0.22	-0.01	-0.05	0.27	-0.06	0.06	0.08
## Easting	-0.22	1.00	-0.27	-0.07	-0.29	-0.05	-0.10	-0.11
## Northing	-0.01	-0.27	1.00	0.08	0.52	0.15	0.26	0.18
## PHI	-0.05	-0.07	0.08	1.00	0.08	-0.05	0.00	0.04
## Buffer3	0.27	-0.29	0.52	0.08	1.00	-0.05	0.26	0.32
## no_NVC	-0.06	-0.05	0.15	-0.05	-0.05	1.00	0.08	-0.24
## no_MSG	0.06	-0.10	0.26	0.00	0.26	0.08	1.00	0.05
## meandbh	0.08	-0.11	0.18	0.04	0.32	-0.24	0.05	1.00
## meanph	-0.21	0.27	-0.14	-0.01	-0.36	0.08	-0.37	-0.26
## meanSOM	0.24	-0.42	0.19	0.13	0.29	-0.04	0.25	0.06
## area_ratio	0.10	0.09	-0.10	0.01	0.17	-0.10	0.20	-0.02

##	meanph	meanSOM	area_ratio
## Alt_m	-0.21	0.24	0.10
## Easting	0.27	-0.42	0.09
## Northing	-0.14	0.19	-0.10
## PHI	-0.01	0.13	0.01
## Buffer3	-0.36	0.29	0.17
## no_NVC	0.08	-0.04	-0.10
## no_MSG	-0.37	0.25	0.20
## meandbh	-0.26	0.06	-0.02
## meanph	1.00	-0.18	-0.10
## meanSOM	-0.18	1.00	0.20
## area_ratio	-0.10	0.20	1.00

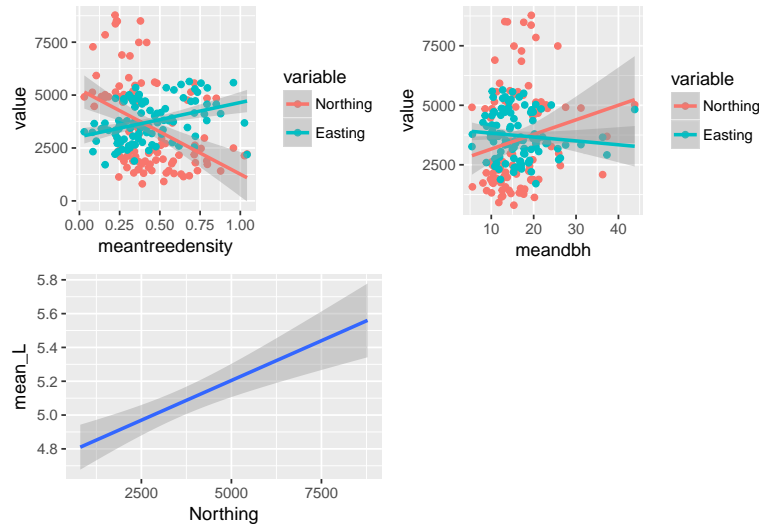
Correlations over 0.3: Buffer and Northing 0.52 meandbh and buffer 0.32 meanph and buffer -0.36 meanph and no_MSG -0.37 meanSOM and Easting -0.42

There is a set of correlations with Northing to buffer to meanph to no MSG which look potentially confounded.



The northern woods tend to have larger buffers, a lower ph and are more mature. There is no reason for meandbh , buffer or meanpH to correlate with each other, so it seems likely that the correlation is related through the latitude to the distribution and age class of the trees and the pH of the woodland.

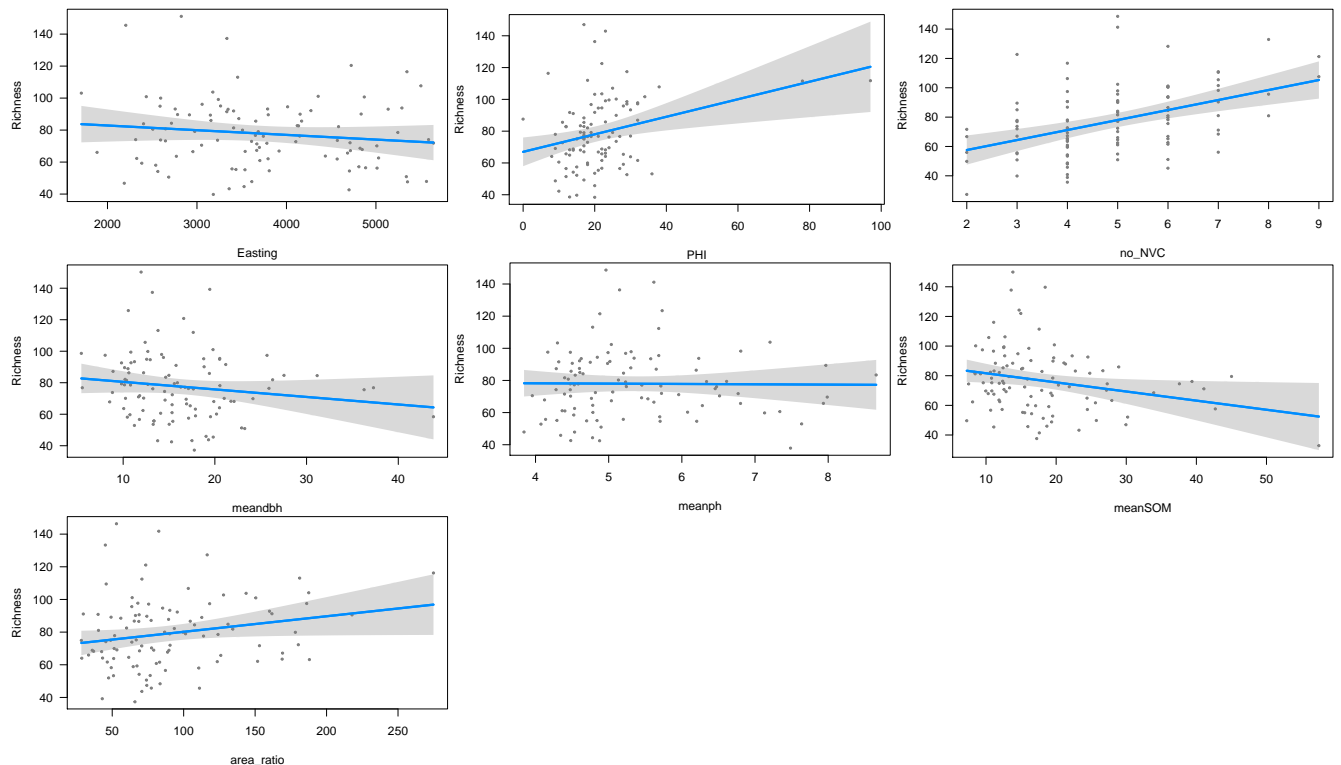
There is also the correlation of richness to buffer, which may have ecological reasons, does the buffer provide a seed bank for recolonization or protect the woodland from nitrification. But I think if you look at the density of the woods, then the relationship is explained.



The northern woods are both less dense, and have trees with greater dbh and more light loving plants. This suggests that there is an increase in richness due to the reduced tree density which allows the addition of light loving species. It is probably not the buffer or the latitude that is contributing to the richness, but the openness of the woodlands.

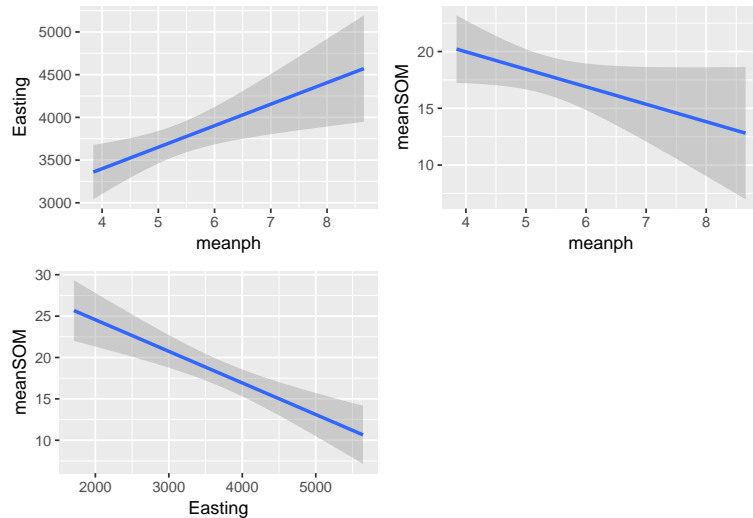
I have removed altitude and number of major soil groups because they dont correlate with richness. Although meanpH does not appear to correlate either, I am not removing it yet, an there ecological reasons why you would expect it to, and perhaps the relationship is affected by the other variables. I will also remove buffer and Northing, because I suspect the correlation with richness is related to the density of the woodlands or pH, not the size of the buffer.

```
## [1] "Easting"      "PHI"          "Richness"     "no_NVC"       "meandbh"
## [6] "meanph"      "meanSOM"      "area_ratio"
```



```
##           Easting  PHI no_NVC meandbh meanph meanSOM area_ratio
## Easting      1.00 -0.07 -0.05 -0.11  0.27  -0.42      0.09
## PHI          -0.07  1.00 -0.05  0.04 -0.01   0.13      0.01
## no_NVC       -0.05 -0.05  1.00 -0.24  0.08  -0.04     -0.10
## meandbh      -0.11  0.04 -0.24  1.00 -0.26  0.06     -0.02
## meanph        0.27 -0.01  0.08 -0.26  1.00  -0.18     -0.10
## meanSOM       -0.42  0.13 -0.04  0.06 -0.18  1.00      0.20
## area_ratio    0.09  0.01 -0.10 -0.02 -0.10  0.20      1.00
```

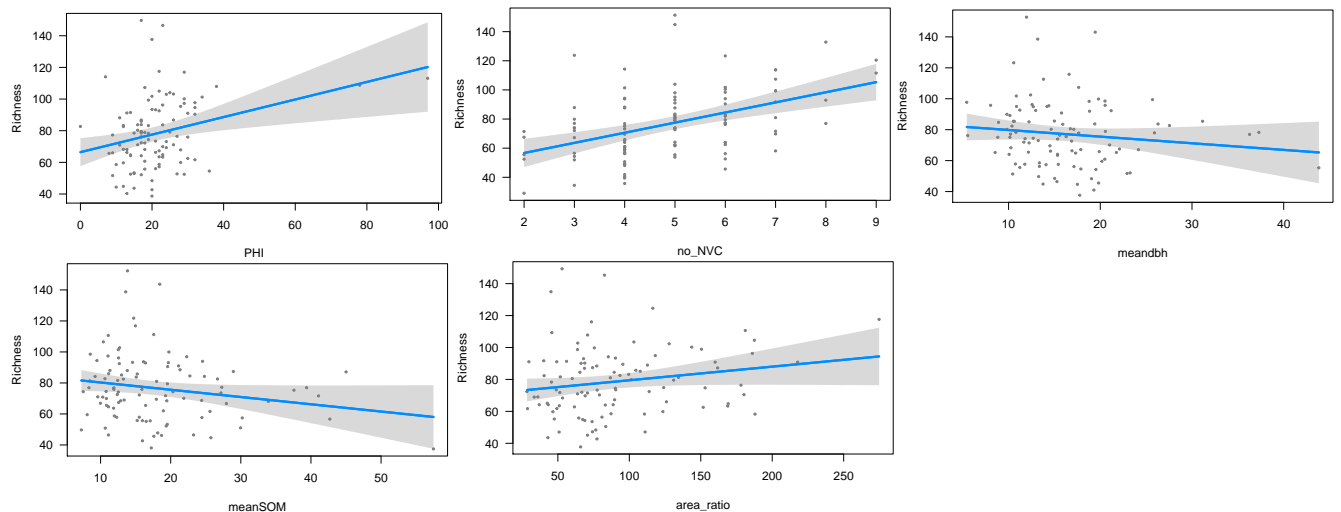
There is a correlation with meanSOM and Easting which we saw previously.



As expected, mean pH decreases with mean SOM, and mean SOM in the Eastern woods is less, and the pH is less. The richness is strongly correlated with meanSOM, and this is related to easting and mean pH.

Therefore the mean SOM would be sufficient to explain richness.

```
## [1] "PHI"           "Richness"       "no_NVC"         "meandbh"        "meanSOM"
## [6] "area_ratio"
```



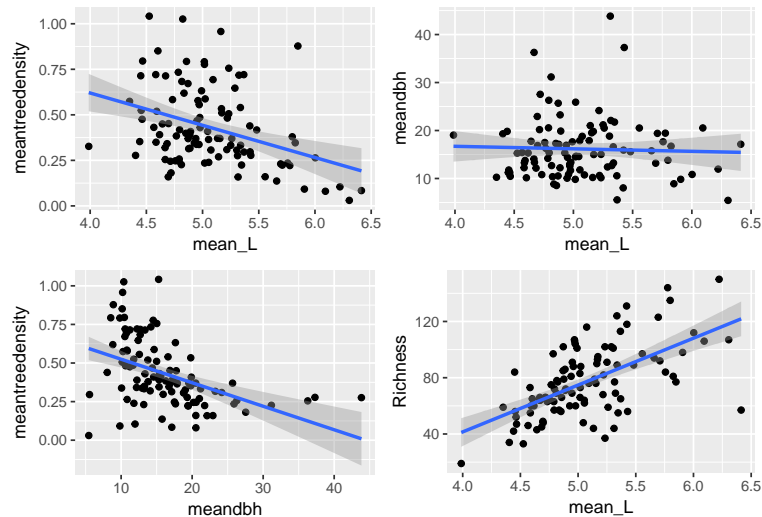
```
##           PHI no_NVC meandbh meanSOM area_ratio
## PHI          1.00 -0.05  0.04   0.13      0.01
## no_NVC       -0.05  1.00 -0.24  -0.04     -0.10
## meandbh       0.04 -0.24  1.00   0.06     -0.02
```

```
## meanSOM      0.13 -0.04  0.06  1.00  0.20
## area_ratio  0.01 -0.10 -0.02  0.20  1.00

##           p  R2
## (Intercept) 0 0.29
## PHI          0  -
## no_NVC       0  -
## meandbh      0.23 -
## meanSOM      0.06 -
## area_ratio   0.07 -
```

meandbh is not significant, and we saw above that reduced tree density allowed more light loving plants. This is not exactly the same as meandbh. You can have lots of large trees, or few large trees and have the same meandbh. So I will add treedensity back in.

```
##           p  R2
## (Intercept) 0 0.31
## PHI          0.01 -
## no_NVC       0  -
## meandbh      0.08 -
## meanSOM      0.04 -
## meantreedensity 0.14 -
## area_ratio   0.09 -
```



““