AB plant and invert responses to HF

Cari D. Ficken

2019-04-26

The goal of this project is to compare and contrast the responses of plants and invertebrates to environmental gradients. It uses the following data:

1. **Plant occurence data**
   - These are presence-absence data from ABMI monitoring sites sampled with the Terrestrial and Wetland protocols
   - Sites are restricted to those classified as wetlands (bogs, fens, marshes, SOWWs, wet meadows)
2. **Insect abundance data**
   - These are frequency data from ABMI monitoring sites sampled with the *Wetland protocol only.*
   - Insects were not identified at ABMI sites sampled with the Terrestrial protocol
3. **Climatic data**
   - These include standard measure of temperature and precip (e.g. evaporation, frost-free period, MAP, MAT. . .)
   - These are available from ABMI website
   - We currently only have data for one year (**which?**) and for sites sampled with the wetland protocol
4. **Human footprint data (HF)**
   - 250 m buffer around each site

1. Load vegegtation and HF datasets

Keep only data from sites with data in both df's.

The vegetation file (`veg_pa`) includes vegetation occurrence (i.e. presence-absence) for all wetland ABMI sites in Alberta. The vegetation taxa are ID'd to species - we have excluded taxa ID'd to genus and we have also removed any finer-level classifications (e.g. subsp.).

The human footprint df (`hf`) provides area estimates (km^2) of different specific human footprint types. Each type (`FEATURE_TY`) has been assigned a higher-order category (`HFCateogry`) based on ABMI's HF metadata. Sum the total disturbance of each HF Category at each site in `hf`.

Let's explore the distribution of sites. We have wetlands in total.

```
## # A tibble: 2 x 2
##   Protocol        n
##   <fct>       <int>
## 1 Terrestrial   773
## 2 Wetland      1281
```

```
## # A tibble: 8 x 2
##   WetlandType     n
##   <fct>       <int>
## 1 Alkali Fen      1
## 2 Bog           445
## 3 Marsh         511
## 4 Poor Fen      321
## 5 Rich Fen      124
## 6 Shallow Lake  235
## 7 Swamp          81
```

```
## 8 Wet Meadow      336
```

```
## # A tibble: 6 x 2
##   NRNAME              n
##   <fct>           <int>
## 1 Boreal           1241
## 2 Canadian Shield    59
## 3 Foothills         114
## 4 Grassland         435
## 5 Parkland          176
## 6 Rocky Mountain     29
```

2. Examine the human footprint (HF) data

The plots below show the total land develped in each natural region, as divided by each HF Category. It also shows the proportion of total developed land in each NR that is attributed to each HF category. Note that wetlands in the Canadian Shield do not have any development in their immediate vicitinty.
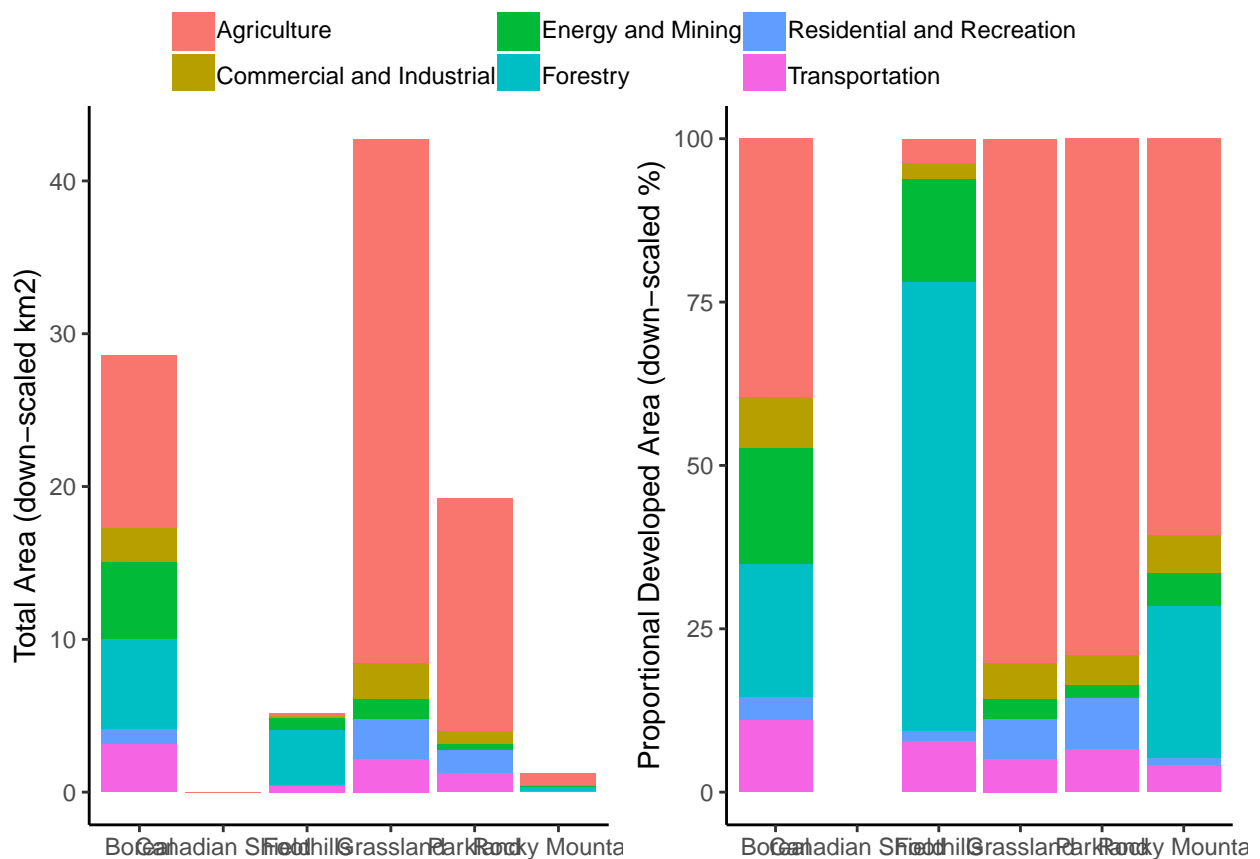
Create a new df with the total human development at each site (`hf_tot`). The histogram below shows the distribution of each HF Category across all focal sites.

Distribution of total human development in AB
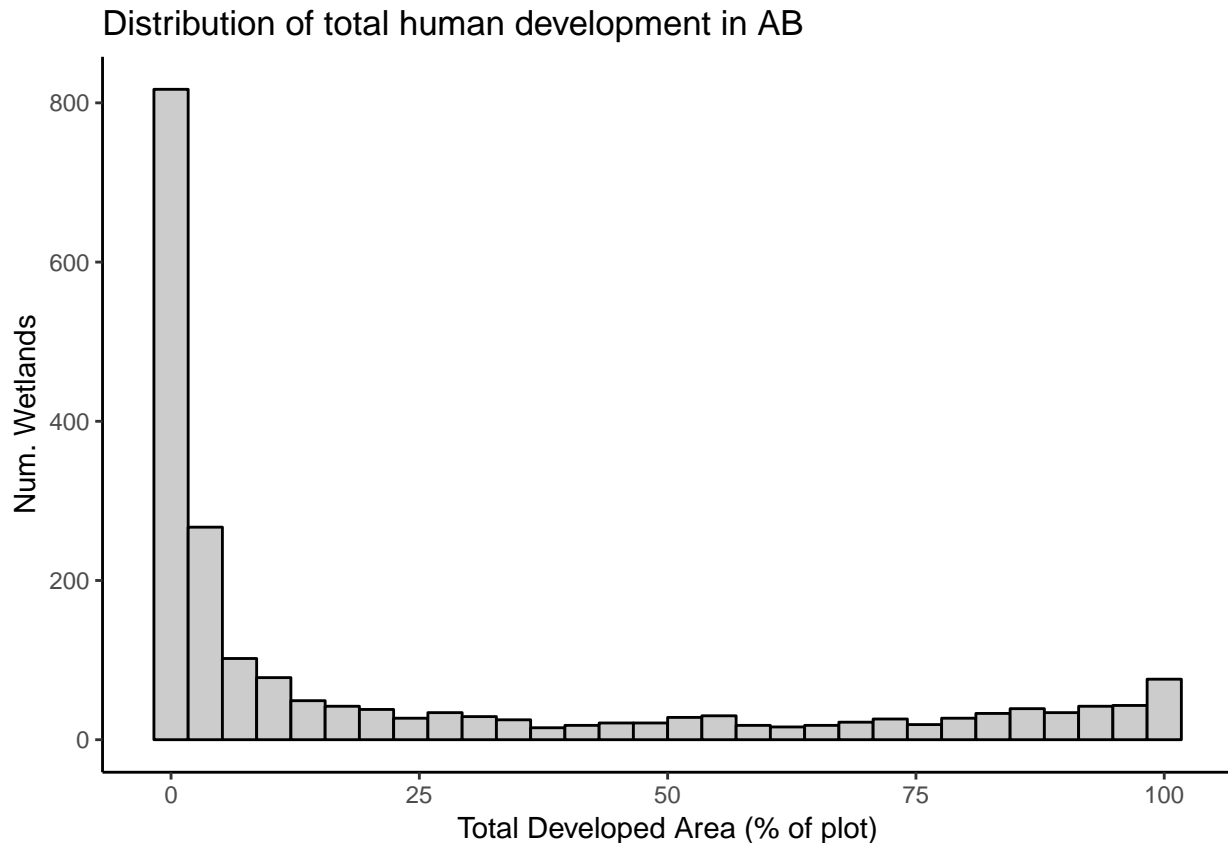
3. Calculate species sensitivity index (SSI)

We can calculate the sensitivity of a species to HF by examining its total number of occurrences in each climate bin. To do so, we:

1. Bin the continuous HG gradient into 10 bins each with a similar number of sitees (~140 sites) and assign each site to a bin
2. Join the vegetation df (`veg_pa`) to the HF df (`hf_tot`) to create `veg_hf`
3. Sum the number of occurrencs of each species in each bin (`occ_freq`)
4. Calculate the coefficient of variation (CV) (sd/mean) for the `occ_freq` of each species across the full HF bin gradient. For example, below we can see the occurence frequency distribution of *Typha latifolia* across the HF gradient.

- Exclude specices which occur only 1x in the full data set since these species will have very high sensitivity (N=34 sp which occur 1x)
- Also be sure to expand the HF bin gradient for every species; that is we have to fill in `occ_freq` = 0 in some bins

4. The resulting df is called `sp_SSI`

3.1 Relationships between species occurrence frequency and CV:

The resulting df is called `sp_SSI`. We can examine the relationship between SSI and occurence frequency; we would expect common species to have low sensitivity to HF (low SSI).
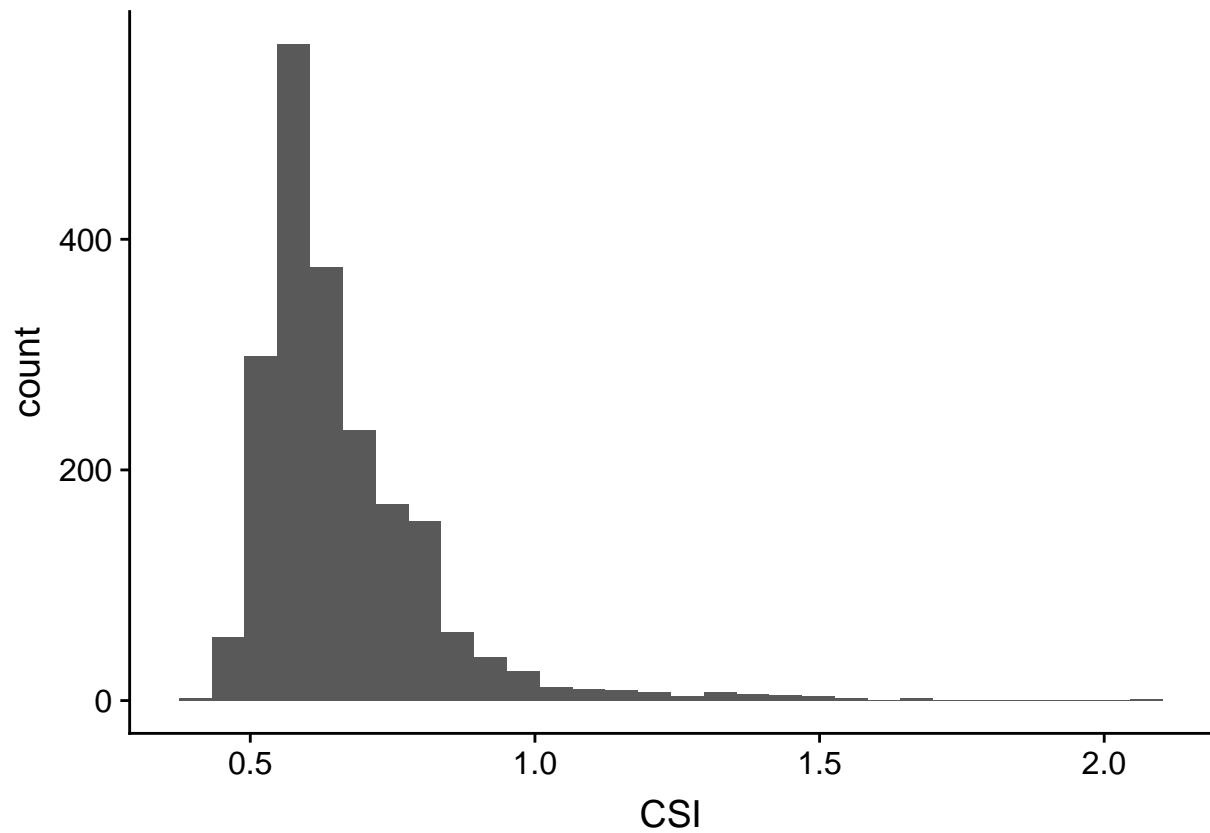
```
## # A tibble: 6 x 2
```

```
##    Species               CV
##    <fct>                <dbl>
## 1 Abies balsamea        0.558
## 2 Abies bifolia         1.09
## 3 Acer negundo          1.94
## 4 Achillea alpina       0.705
## 5 Achillea millefolium  0.306
## 6 Acorus americanus     1.40
```

Human Footprint
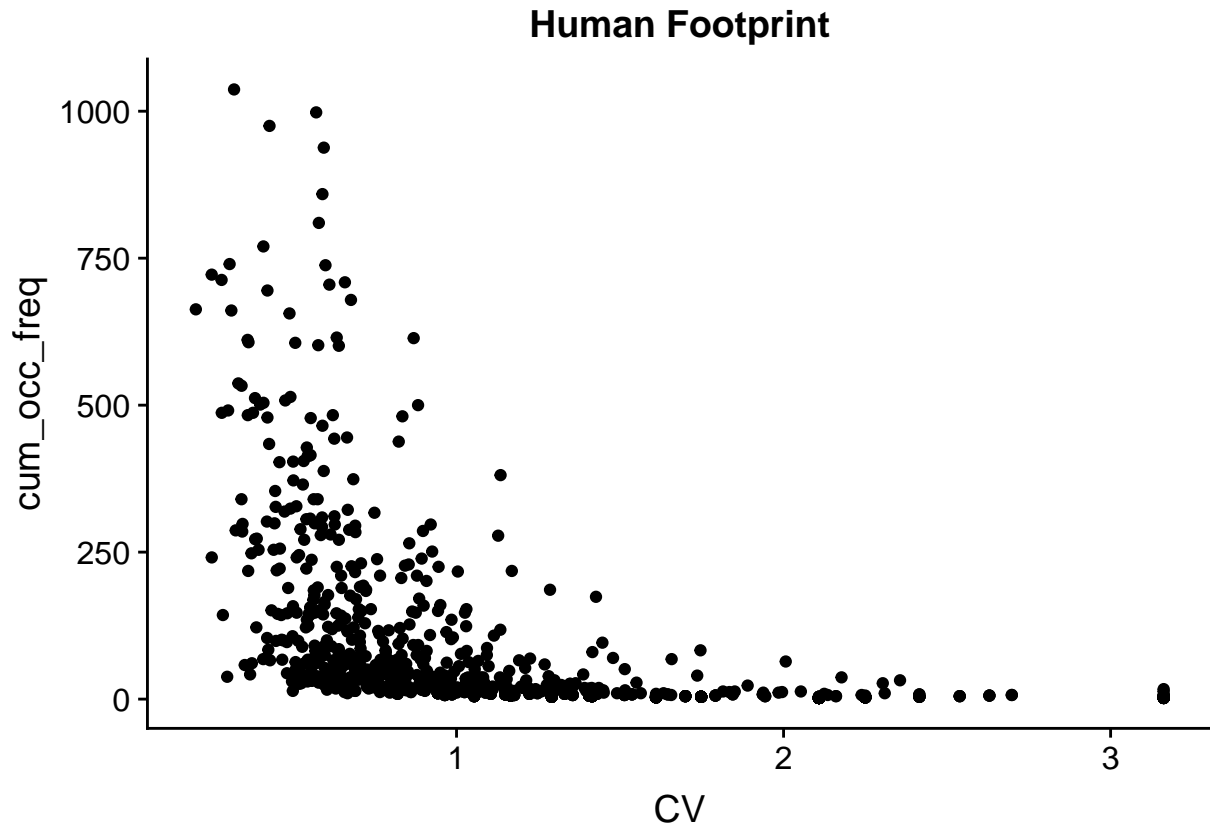
# 4. Calculate community sensitivity index (CSI)

Add SSI of each species (`sp_SSI`) to the vegetation df (`veg_pa`). Calculate the mean SSI at each site (`CSI`).

5. Examine patterns of CSI across different wetland/site classifications gradients

5.1 Overall patterns

Random effects models (linear, 2nd order polynomial, 3rd order polynomial) show no effect of including a Protocol x Site unique ID random effect. Linear models find that the quadratic model describes CSI better than the linear model.

Poly: R2 = 0.28, AIC = −2408, p < 0.001
Linear: R2 = 0.24, AIC = −2301, p < 0.001
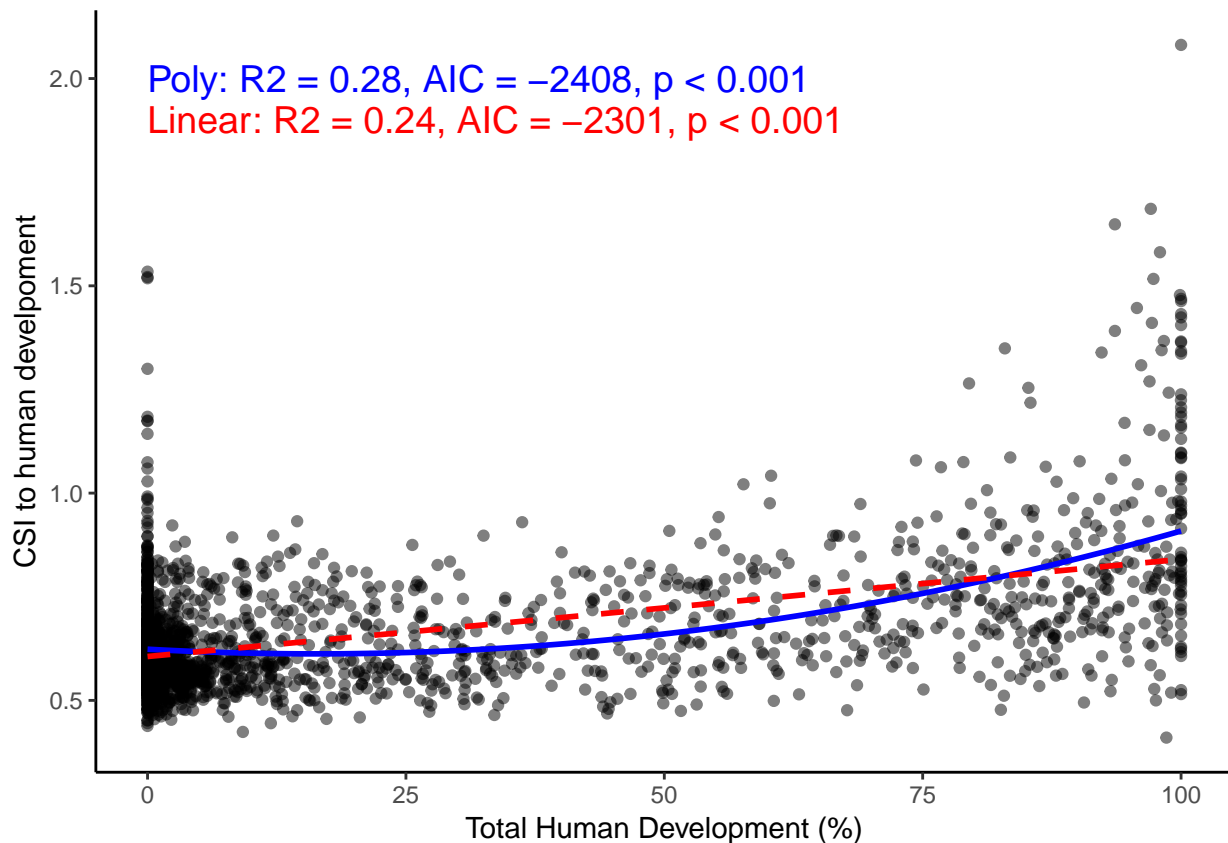
5.2 Comparisons between Terrestrial and Wetland Protocols

Compare the distribution of CSI values for sites sampled with wetland vs terrestrial ABMI protocols.

The Wetland protocol underestimates the CSI of communities at high development levels. Generally, though, communities show the highest sensitivity to development at really high development intensities. This may be because communities at high develpoment levels are composed of ruderal species that are only found under these disturbed conditions; at low development intensitites, perhaps communities are composed of more broadly distributed native species. That is, after 25% area is development, ruderals take over.

The variance explained by including Protocol as a random effect was again indistinguishable from zero. However, Protocol was a significant predictor in quadratic models, and including it as an interaction term significantly improved model fit (R2=0.35; AIC=-2614) compared to models without Protocol.

```
## 
## Call:
## lm(formula = CSI ~ poly(totdist_percent, 2) * Protocol, data = veg_CSI_HF)
## 
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.54317 -0.07119 -0.02135  0.05506  1.01325
## 
## Coefficients:
##                                 Estimate Std. Error t value
## (Intercept)                     0.687462   0.004679 146.920
## poly(totdist_percent, 2)1       4.980429   0.207006  24.059
## poly(totdist_percent, 2)2       2.730414   0.216192  12.630
```

6

```
## ProtocolWetland                          -0.043448   0.005898  -7.367
## poly(totdist_percent, 2)1:ProtocolWetland -2.350379   0.263759  -8.911
## poly(totdist_percent, 2)2:ProtocolWetland -2.363482   0.269588  -8.767
##                                            Pr(>|t|)
## (Intercept)                                < 2e-16 ***
## poly(totdist_percent, 2)1                  < 2e-16 ***
## poly(totdist_percent, 2)2                  < 2e-16 ***
## ProtocolWetland                            2.52e-13 ***
## poly(totdist_percent, 2)1:ProtocolWetland  < 2e-16 ***
## poly(totdist_percent, 2)2:ProtocolWetland  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1278 on 2048 degrees of freedom
## Multiple R-squared:  0.3541, Adjusted R-squared:  0.3525
## F-statistic: 224.6 on 5 and 2048 DF,  p-value: < 2.2e-16

##                  df        AIC
## m.poly            4 -2408.150
## m.poly.protocol   7 -2614.656
```
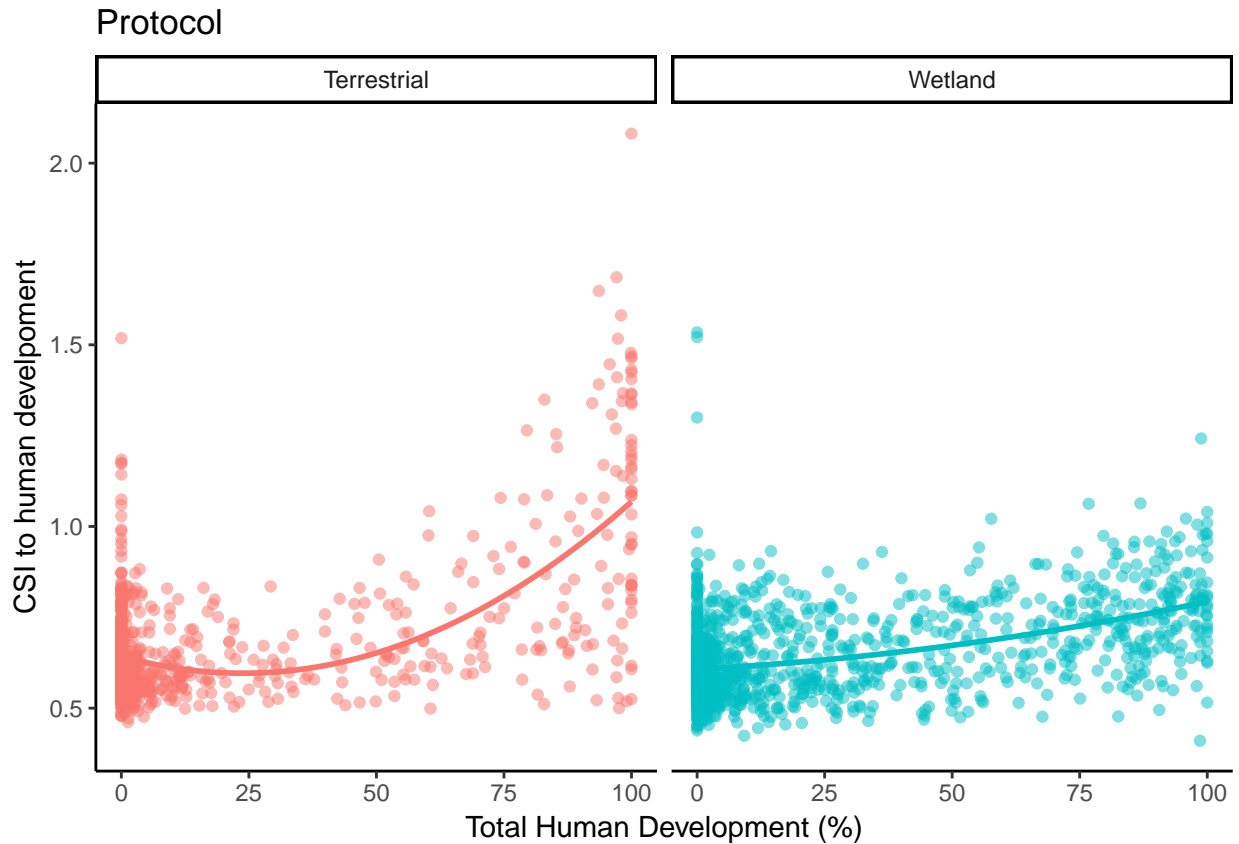

5.3 Comparisons among wetland classes

Patterns across wetland classes are difficult to disentangle due to vastly different sample sizes of each wetland class. Marshes, shallow lakes, and wet meadows have the most thorough representation across the full HF gradient. For these classes, there is slightly high sensitivity to development at both ends of the development

spectrum. This largely parallels the pattern observed across the full dataset. For other wetland classes, sites are not distributed across the gradient sufficiently to really commpare their sensitivity.

No single wetland class appears to have particularly high or low sensitivity overall.
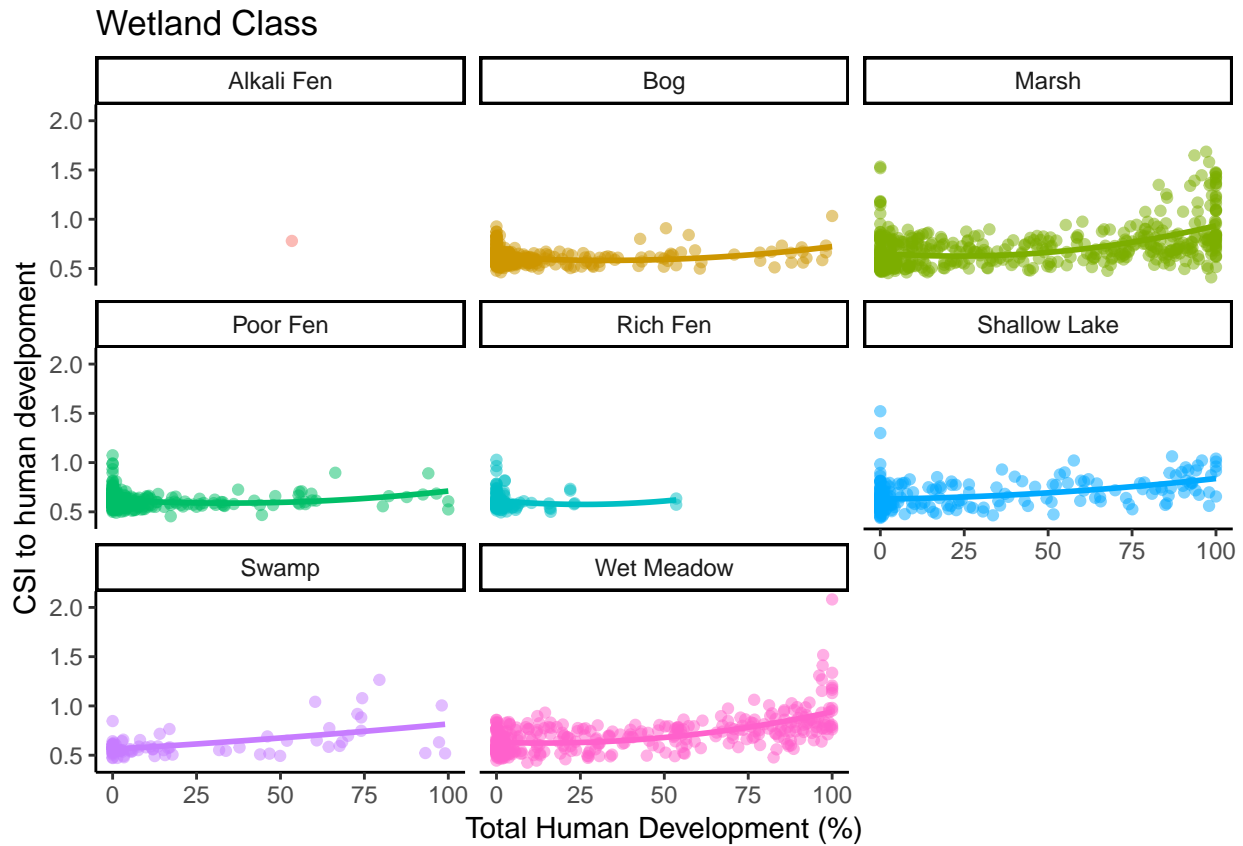
5.4 Comparisons among natural regions

Wetlands across the natural regions vary somewhat in their exposure to human development. For example, there is essentially no development in the Canadian shield and very little in the rocky mountain NR Sites in the boreal, foothills, grassland, and parkland NRs show a greater range in human development.

Rocky mountain wetlands have high sensitivty at low development intensities. Grassland wetlands maybe have higher overall sensitivity to human devleopment than wetlands in the other NRs.
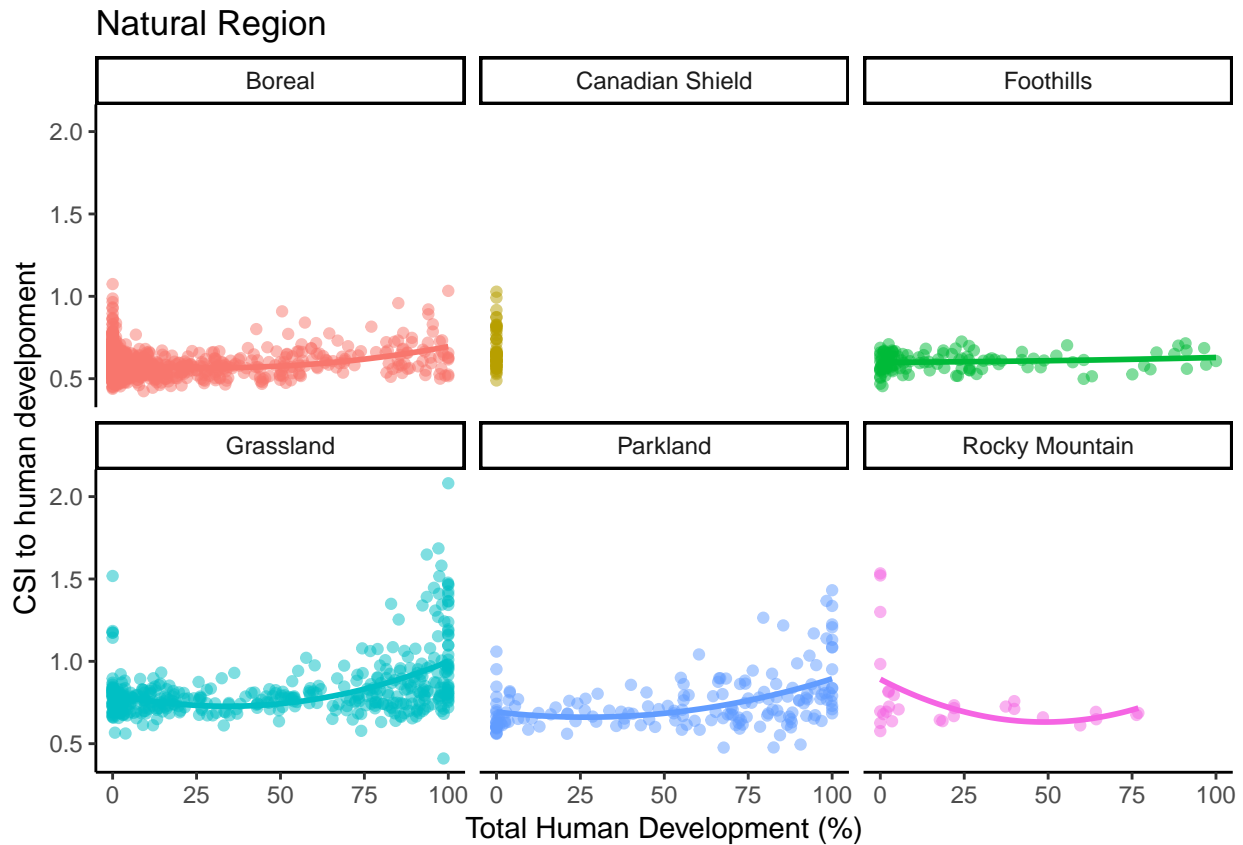
6. Patterns of species richness across HF gradient

Channeling the Intermediate Disturbance Hypothesis here...

Species richness is somewhat elevated at intermediate human development intensitities. This supports the findings of Mayor et al 2012 (*Regional boreal biodiversity peaks at intermediate human disturbance*) which uses ABMI data to examine this pattern only in the boreal.

Variance explained by including unique site ID (i.e. Protocol x Site) as a random effect is greater than zero, indicating this is an important variable to keep. Including Protocol also as a direct effect improves model fit somewhat (AIC w/ Protocol = 16652; AIC w/o Protocol = 16978) you also have an increase in dfs (5 vs 8).

```
##                             df      AIC
## m.poly.raneff.rich           5 16977.51
## m.poly.raneff.rich.protocol  8 16652.38
```
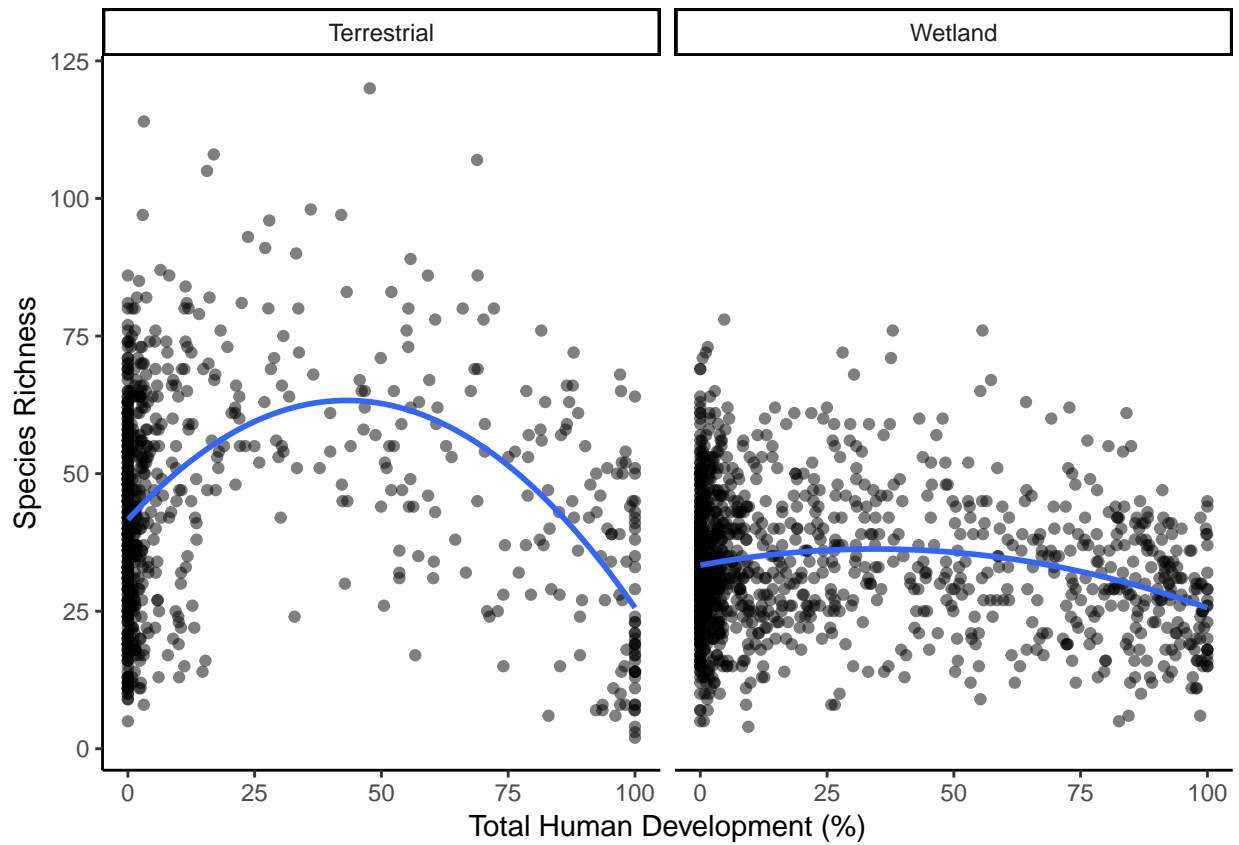
7. Examine patterns of CSI to HF across climatic gradients

Now we can examine whether CSI to disturbance is mediated by mean climatic conditions. First check to ensure that HF isn't correlated with climatic conditions: it definitely is not.
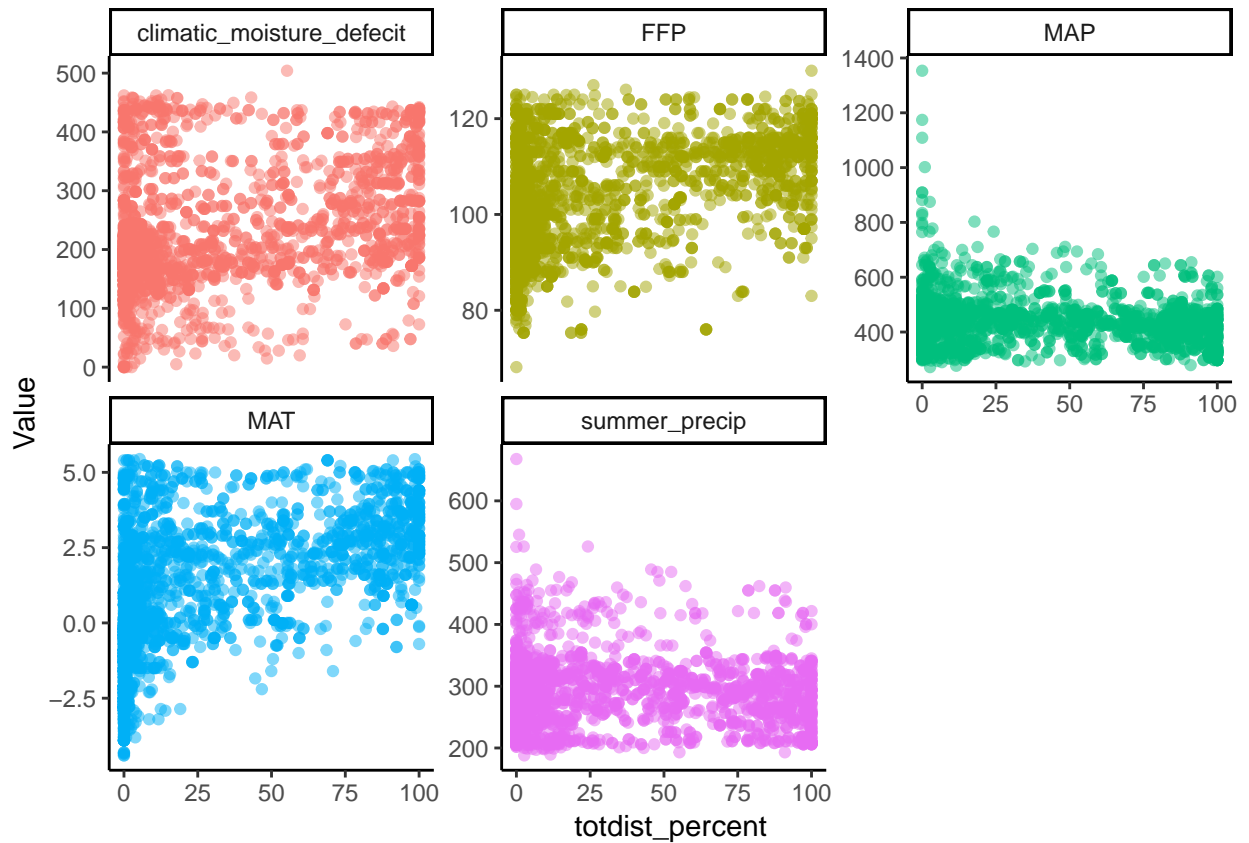
Next we can compare patterns of CSI and climatic conditoins. There is an insufficient gradient in MAP or summer precip to really compare CSI across these variables. There are relationships between CSI and CMD (climatic moisture defecit), and with MAT: communities are relatively more sensitive to human development at either either ends of each climatic gradient. There is a general positive relationship between CSI and CMD, FFP, and MAT, indicating that communities are more sensitive to development in warmer, drier areas. This has important implications for climate change.
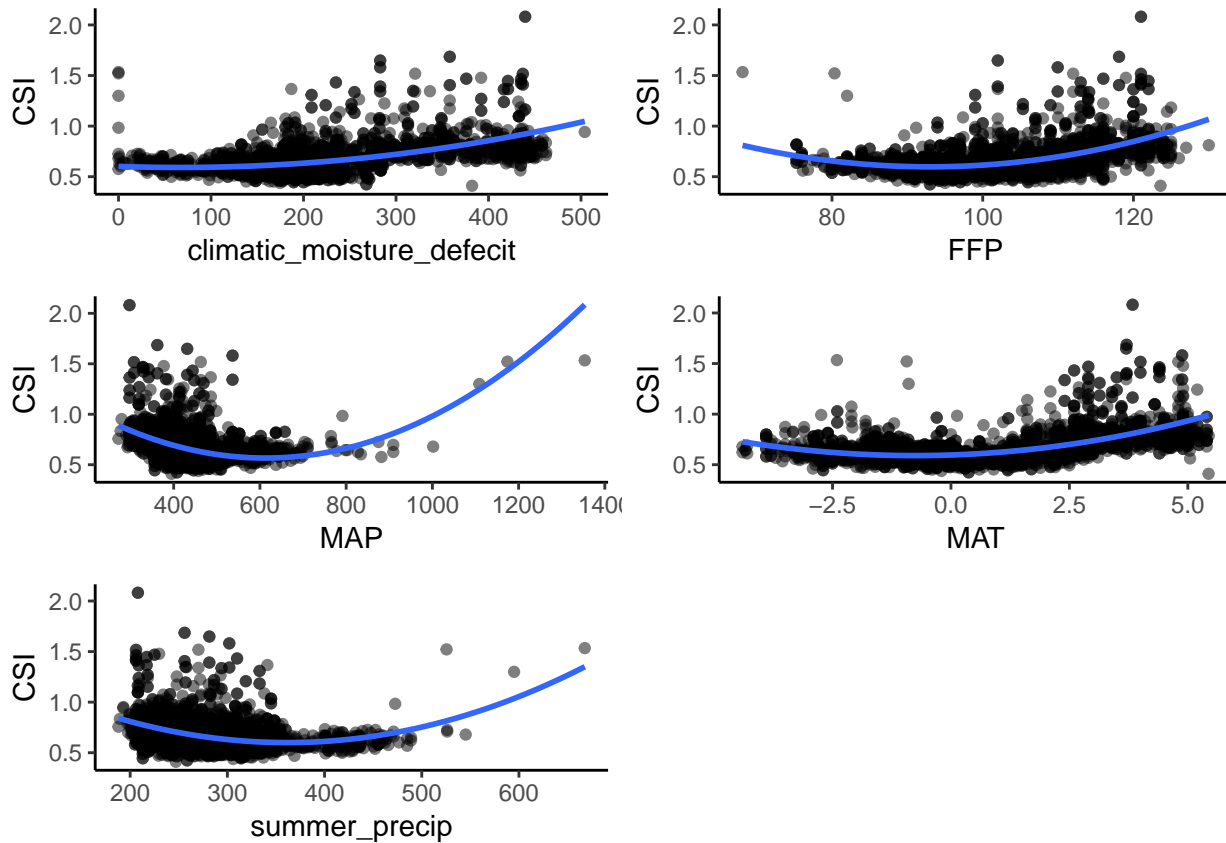
9. Ordination of communities at high and low end of HF gradient

Communities at low and high end of HF gradient show similar plant diversity (richness), but are these communities compositionally different? Yes, MRPP test shows significant difference between these groups and the ordination visualizes this.

Species most strongly associated with sites in the low HF bin (bin 1):

- *Rhodondendron tomentosum*: marsh labrador tea
- *Eriophorum scheuchzeri*: white cotton grass
- *Carex oligosperma*: fewseed sedge
- *Drosera anglica*: English sundew
- _Scheuchzeria palustris: Rannoch-rush

Species most strongly associated with sites in the highest HF bin (bin 10):

- *Euphorbia serpyllifolia*: thymeleaf sandmat
- *Avena sativa*: common oat
- *Polygonum ramosissimum*: bushy knotweed
- *Potentilla supina*: bushy cinquefoil
- *Bromus commutatus*: meadow brome or hairy chess

```
## 
## Call:
## mrpp(dat = veg_d, grouping = as.factor(veg_hf2$HFbin))
## 
## Dissimilarity index: binary jaccard
```

```
## Weights for groups:  n
##
## Class means and counts:
##
##       1      10
## delta 0.8222 0.8883
## n      206    205
##
## Chance corrected within-group agreement A: 0.06089
## Based on observed delta 0.8552 and expected delta 0.9106
##
## Significance of delta: 0.001
## Permutation: free
## Number of permutations: 999
```
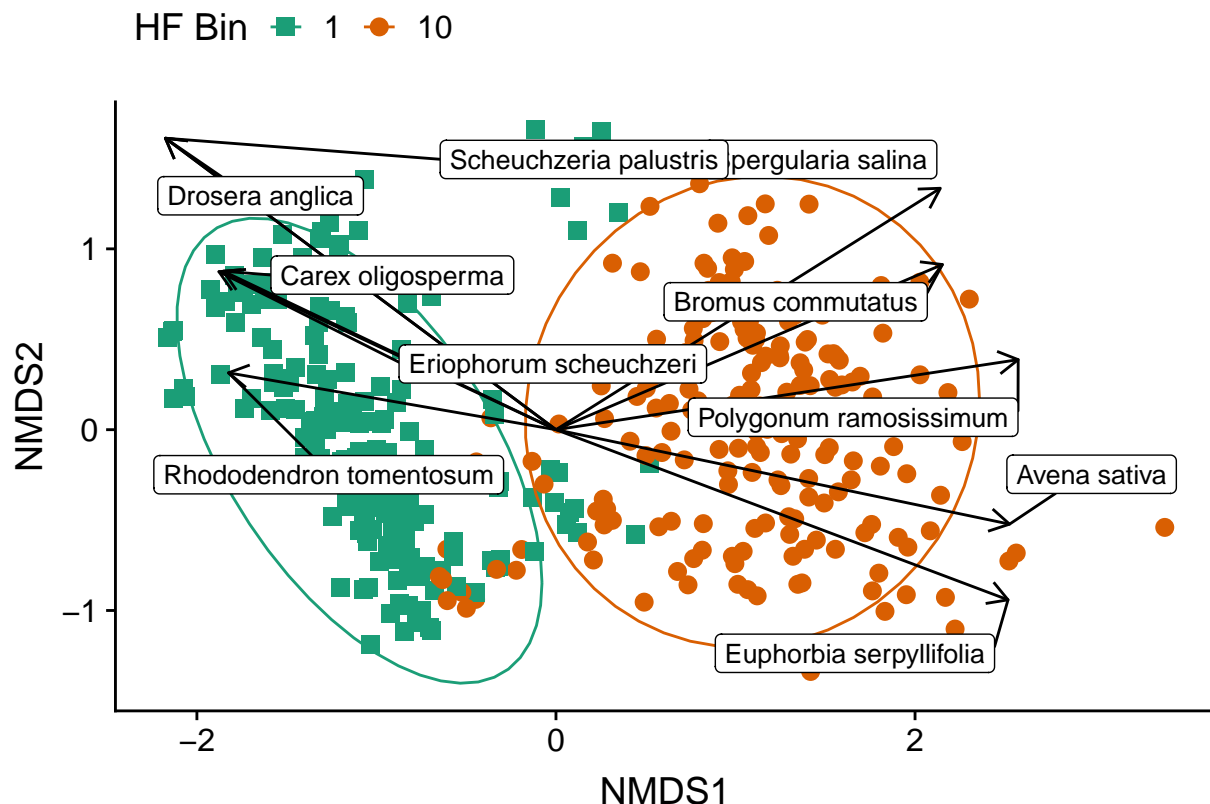

8. Choosing the most predictive variables

How can we chose the most appropriate variablesl to describe community sensitivty? Use regression trees and/or random forest. To predict the community sensitivity index (CSI), use the following predictors

- Protocol (terrestrial vs wetland)
- Natural region
- Wetland class
- Site-level species richness
- Climate variables

Regression tree predicting CSI - MAT
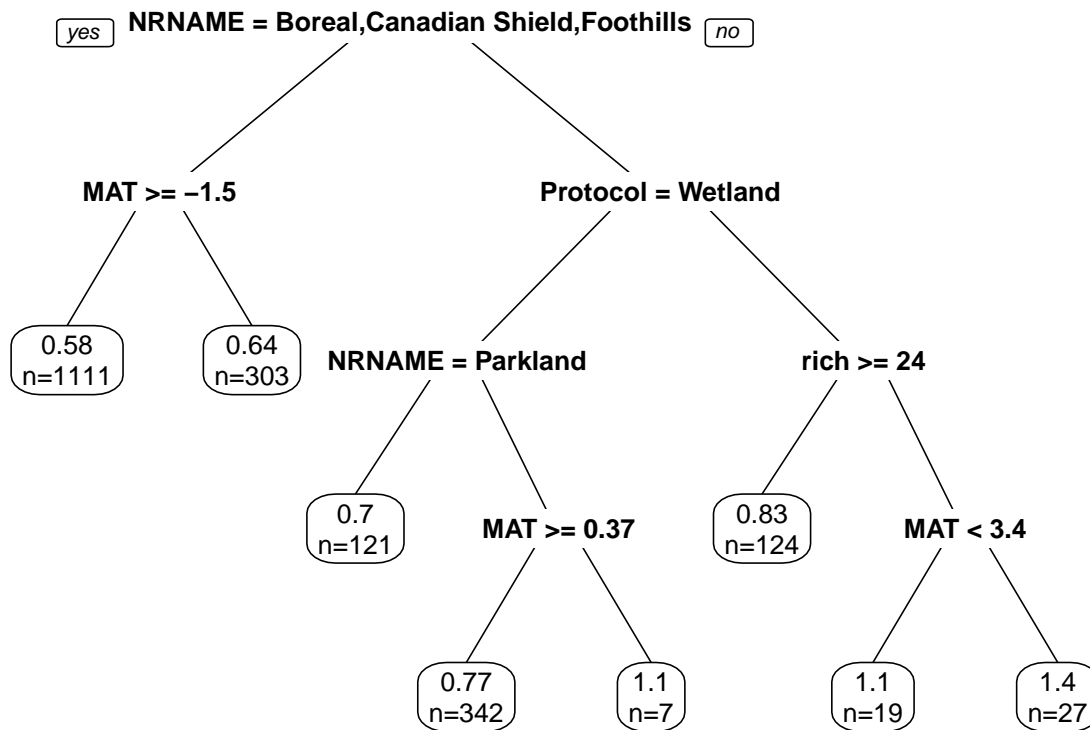
Now grow a random forest to compare the importance of each predictor.

```
##                          IncNodePurity
## Protocol                      5.225763
## NRNAME                        8.929444
## WetlandType                   1.495853
## rich                          7.612543
## climatic_moisture_defecit     6.670627
## MAT                           8.796422
## MAP                           3.698358
## FFP                           4.054552
## summer_precip                 3.031345
```

```
## [1] 0.7267376
```

The final tree's pseudo-$R^2 = 0.73$.