Jess Daly univariate hw

1.

read in tree dataset

```
trees <- read.csv('https://raw.githubusercontent.com/dmcglinn/quant_methods/gh-pages/data/treedata_subset.csv')
```

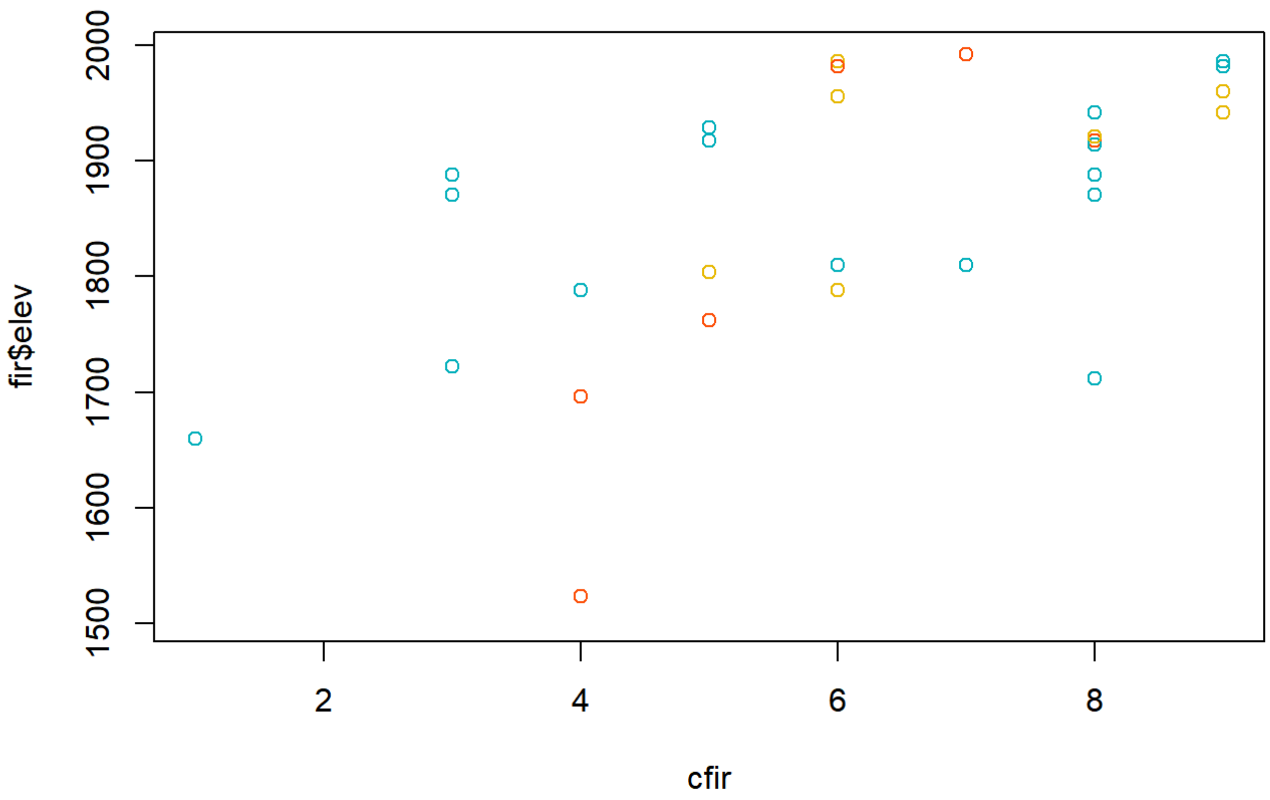
specify I'm only interested in looking at the fir and the red maple, and then create a vector for cover for each type of tree

```
fir <- trees[trees$species == "Abies fraseri", ]
maple <- trees[trees$species == "Acer rubrum", ]
cfir<-fir$cover
cmap<-maple$cover
```

I created plots to invesigate the relationship for each of the non-catigorical variables (elevation, tci, stream distance, and beers) for each tree. Each of the colors represents one of the three disturbance values, which is the only other discrete variable and should also be considered.

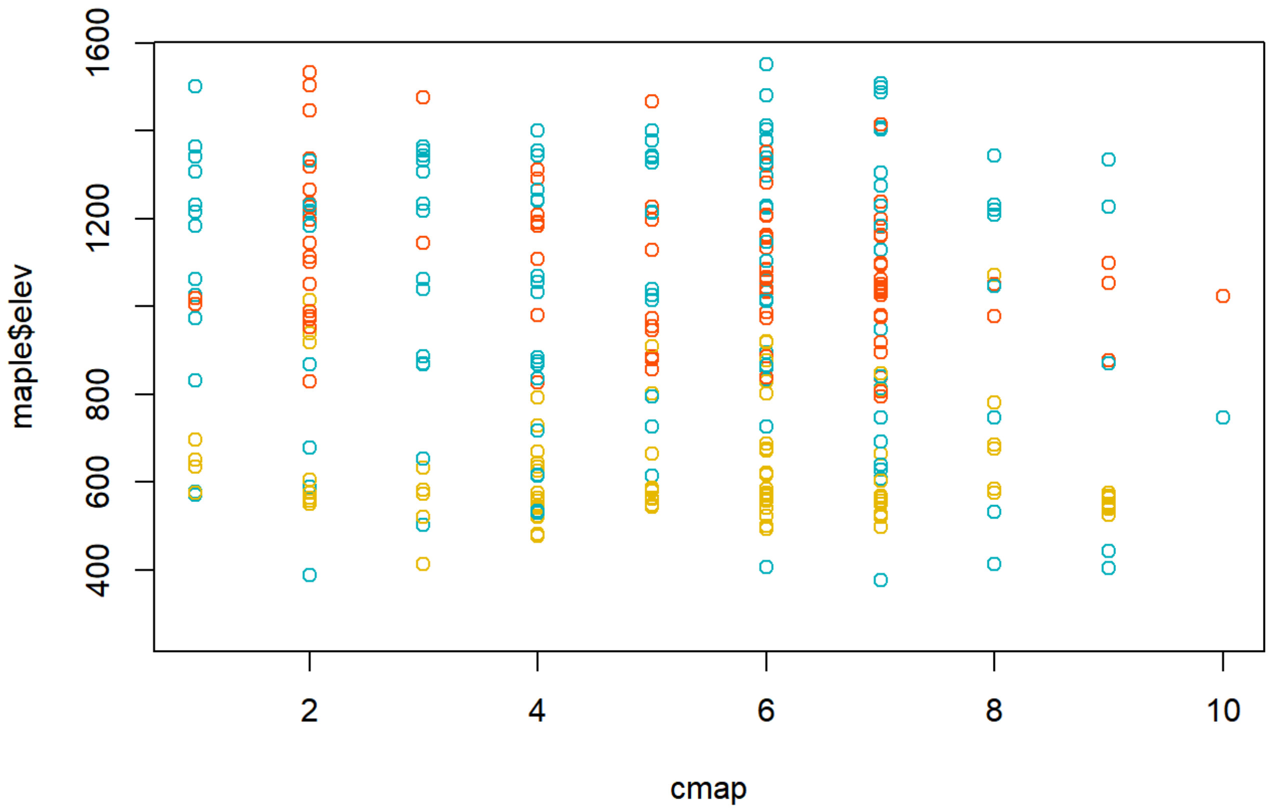
```
colors <- c("#00AFBB", "#E7B800", "#FC4E07")
colors <- colors[as.numeric(fir$disturb)]
colors <- colors[as.numeric(maple$disturb)]
plot(cfir,fir$elev, col=colors)
```

Create PDF in your applications with the Pdfcrowd HTML to PDF API PDFCROWD




```
plot(cmap,maple$elev, col=colors)
```

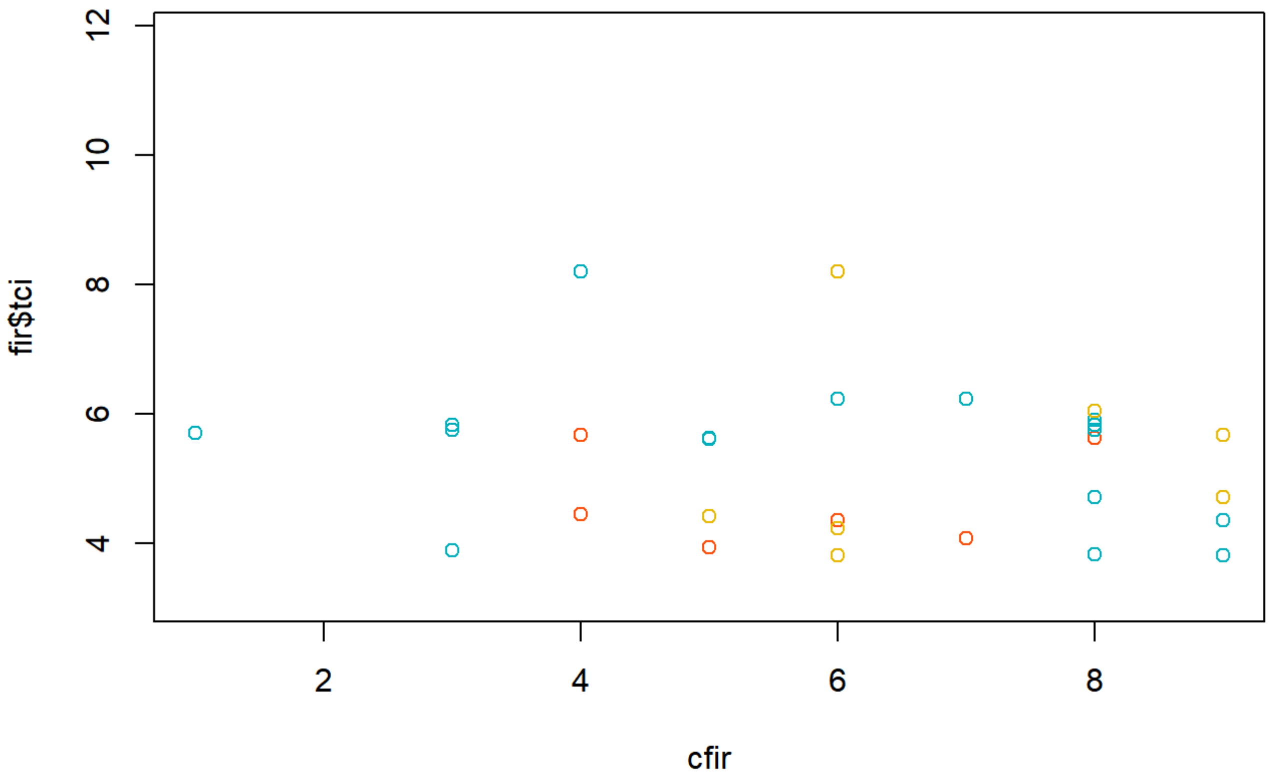
Create PDF in your applications with the Pdfcrowd HTML to PDF API PDFCROWD



From looking at the elevation plots, it seems like maples are more evenly distributed than firs between different elevations. That is, there seems to be a positive relaionship between cover and elevation in the firs, but this does not look to be the case for the maples. However in the maples there does look like there is some kind of relationship between disturbance history and elevation. This would make sense since disturbance history describes logging activity, which would probably be happening in concentrated areas. Additionally from looking at the plots clearly there are more red maples than firs in the forest.

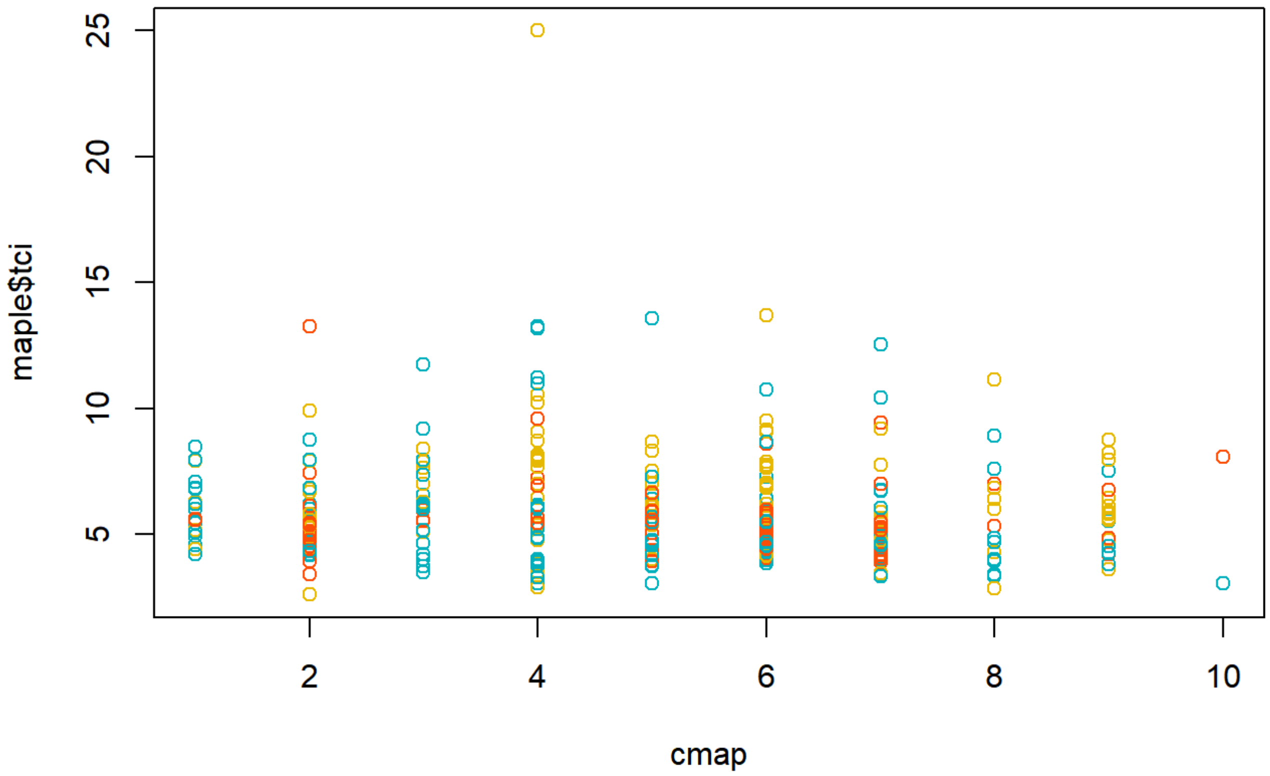
```
plot(cfir,fir$tci, col=colors)
```

Create PDF in your applications with the Pdfcrowd HTML to PDF API PDFCROWD




```
plot(cmap,maple$tci, col=colors)
```

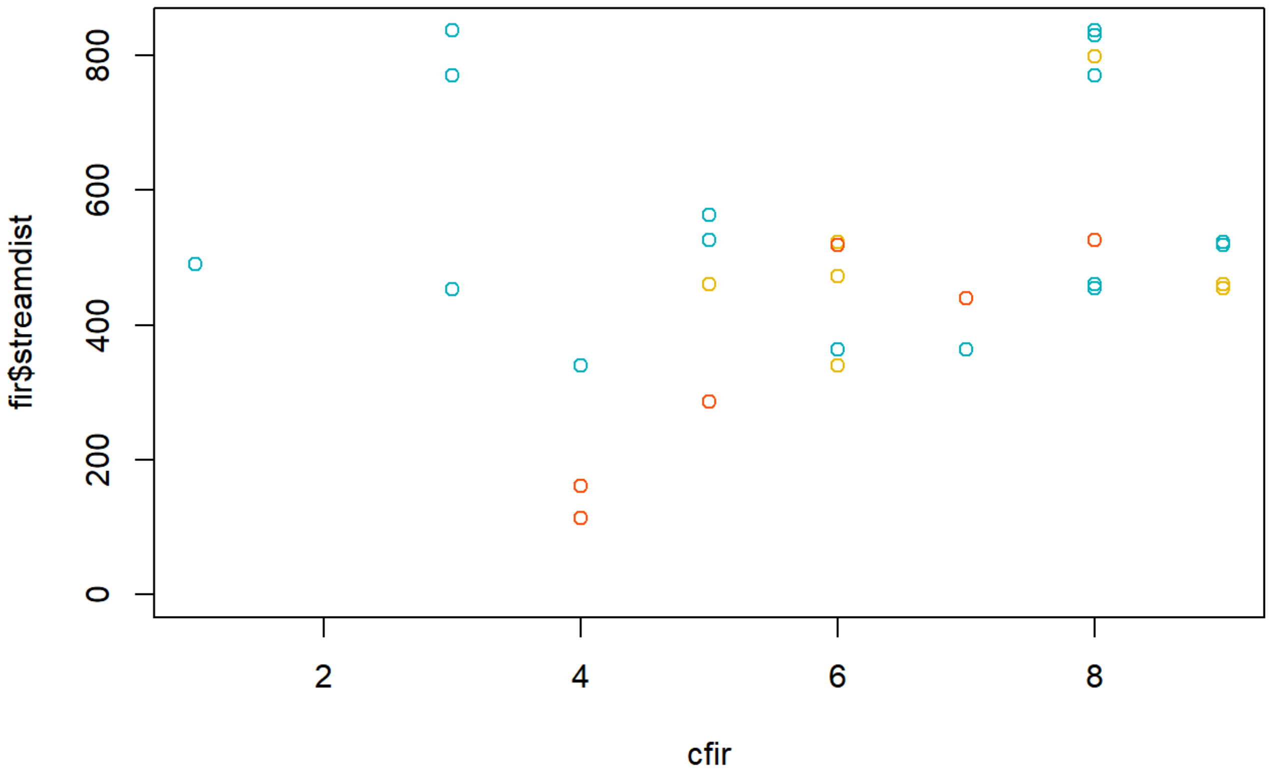
Create PDF in your applications with the Pdfcrowd HTML to PDF API PDFCROWD



There does not appear to be much of a relationship between tci (topographic convergence index) and cover for either species. However from this plot it seems like firs tended to have high covers in the areas which they are present (majorit of points fall above 5, whereas maples look far more evenly distributed between 0-10).

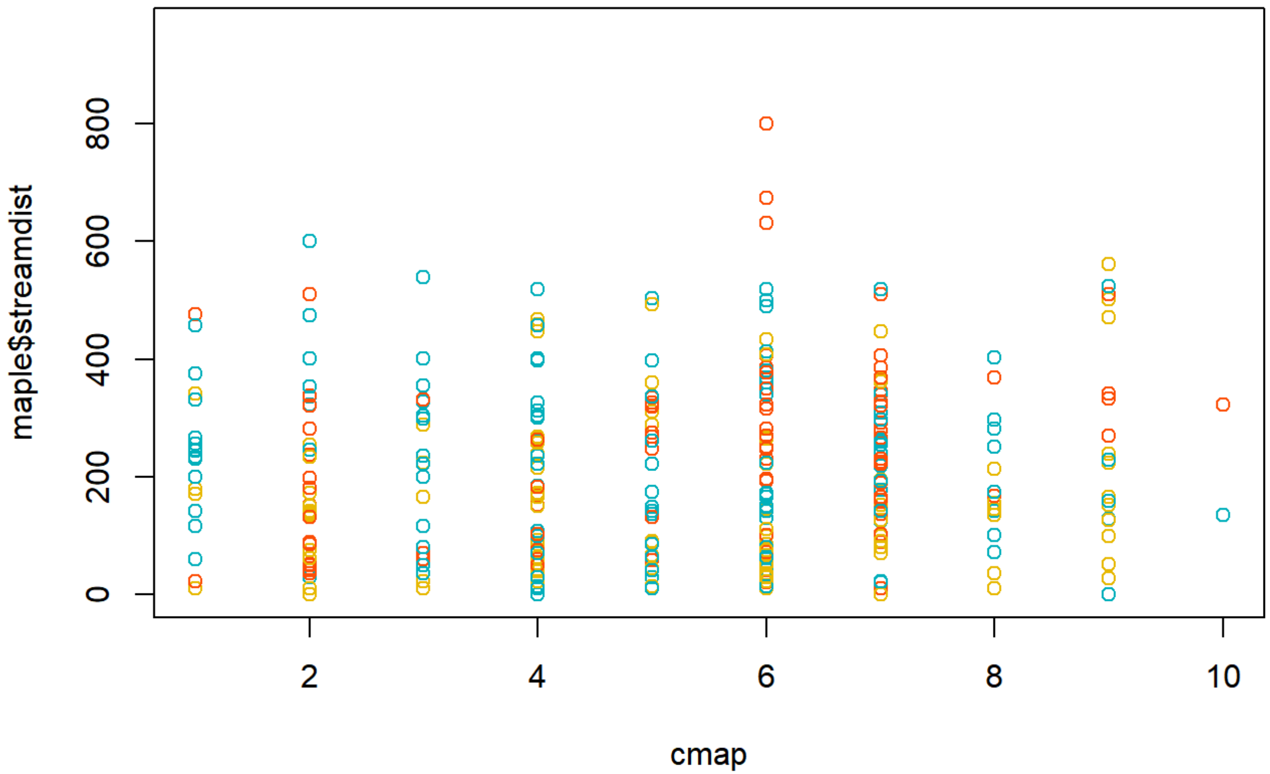
```
plot(cfir,fir$streamdist, col=colors)
```

Create PDF in your applications with the Pdfcrowd HTML to PDF API PDFCROWD




```
plot(cmap,maple$streamdist, col=colors)
```

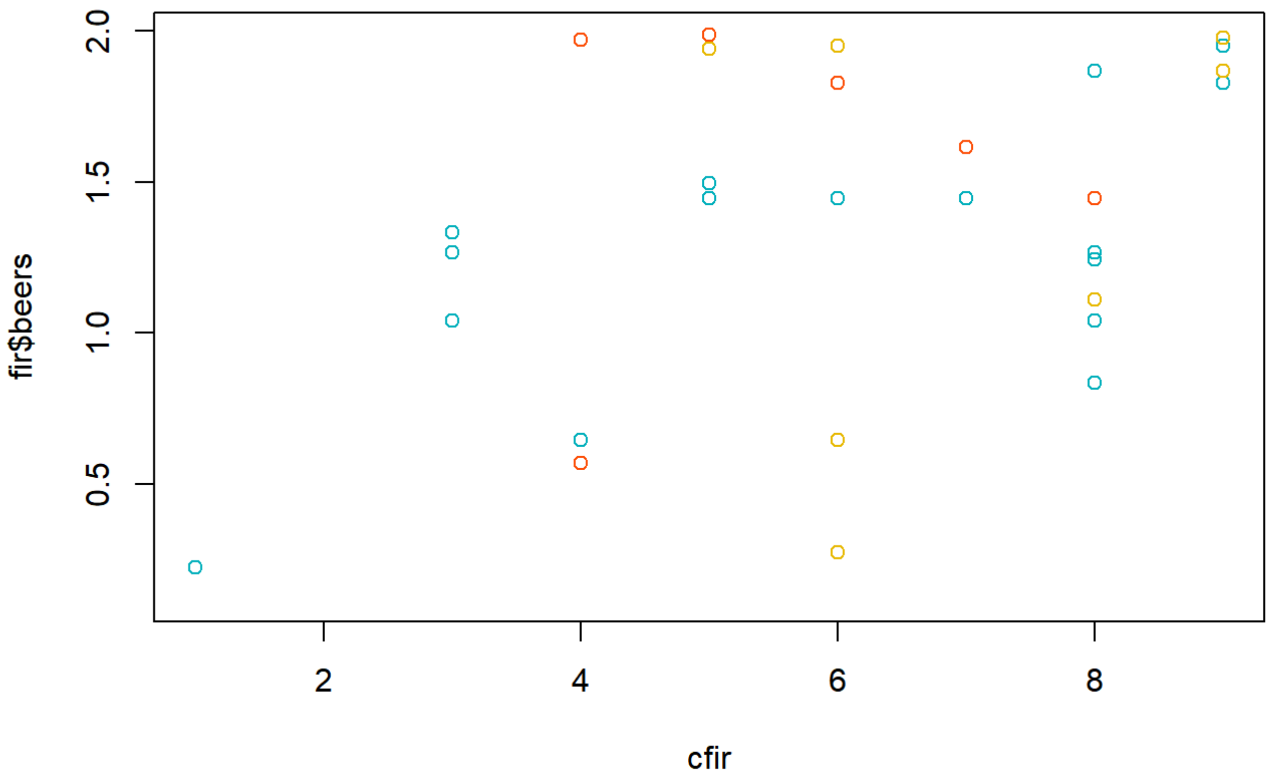
Create PDF in your applications with the Pdfcrowd HTML to PDF API PDFCROWD



Maples do not appear to show a relationship between cover and distance to the nearest stream. Firs look like there may be a vague positive relationship between the two.

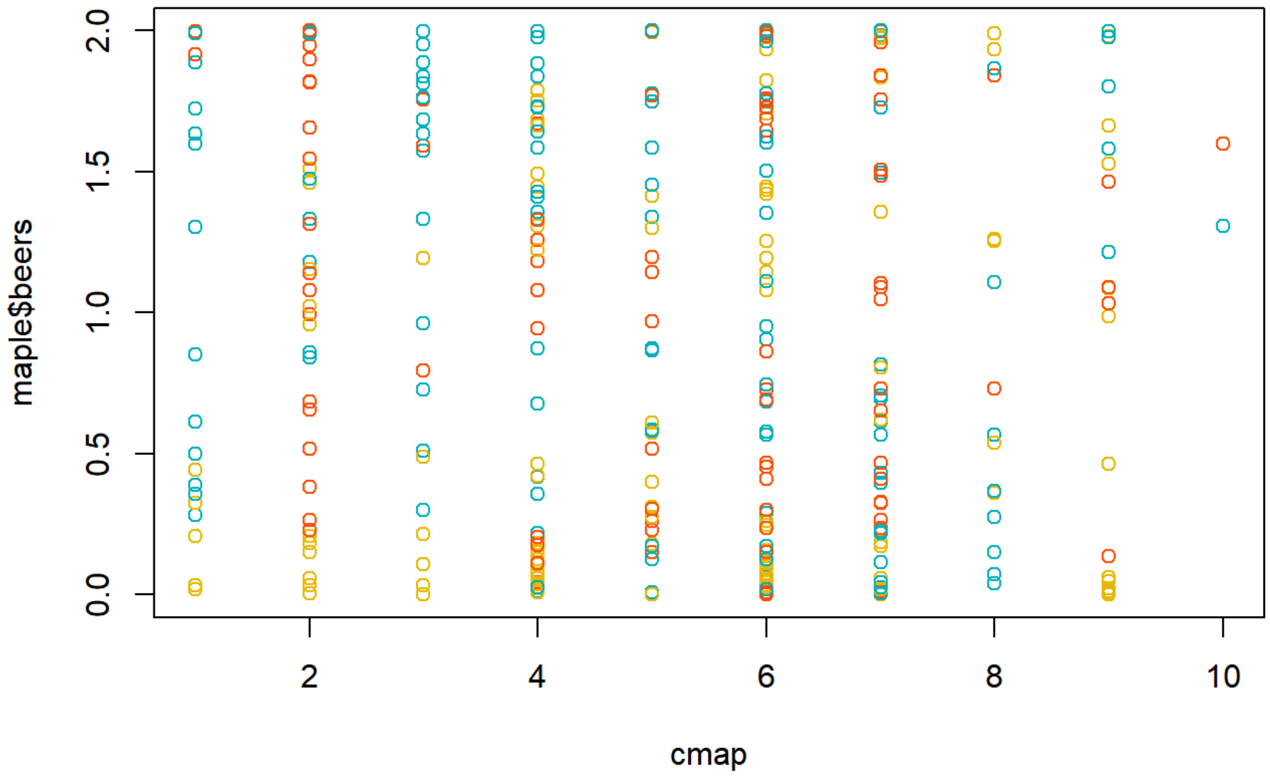
```
plot(cfir,fir$beers, col=colors)
```

Create PDF in your applications with the Pdfcrowd HTML to PDF API PDFCROWD




```
plot(cmap,maple$beers, col=colors)
```

Create PDF in your applications with the Pdfcrowd HTML to PDF API PDFCROWD



Once again maples do not appear to show a relationship between transformed slope aspect and cover, while firs look like they may show a vague positive trend.

In order to gather more information and actually determine if there are relationships between these variables, I will run several models.

First I ran ANOVAS for each type of tree.

```
firanova<-aov(cover~disturb, data=fir)
summary(firanova)
```

Create PDF in your applications with the Pdfcrowd HTML to PDF API PDFCROWD

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## disturb      3   46.2  15.401   3.525 0.0234 *
## Residuals   40  174.8   4.369
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
mapanova<-aov(cover~disturb, data=maple)
summary(mapanova)
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## disturb       3   17.1   5.692    1.39  0.245
## Residuals   719 2944.2   4.095
```

ANOVAs indicate there is a significant effect of disturbance on the firs, but not on the maples.

Next I will use a linear model in order to examine the effects of all of the variables. ANOVAs only work with catigorical variables, so a liner model needs to be used here instead because several of the variables are numeric.

```
linmodfir<-lm(cover~elev+tci+streamdist+beers, data=fir)
summary(linmodfir)
```

```
##
## Call:
## lm(formula = cover ~ elev + tci + streamdist + beers, data = fir)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.2971 -1.0339  0.2795  1.1257  4.2595
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -20.576781   4.153587  -4.954 1.45e-05 ***
## elev          0.013772   0.002438   5.649 1.59e-06 ***
## tci           0.330847   0.180651   1.831   0.0747 .
```

Create PDF in your applications with the Pdfcrowd HTML to PDF API PDFCROWD

```
## streamdist   -0.001590   0.001578  -1.007   0.3199
## beers         0.234717   0.489113   0.480   0.6340
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.62 on 39 degrees of freedom
## Multiple R-squared:  0.5367, Adjusted R-squared:  0.4892
## F-statistic: 11.29 on 4 and 39 DF,  p-value: 3.497e-06
```

Elevation is shown to be the only significant variable affecting cover. It should be noted that the p-value for tci is low, and with a greater sample size may have been found to be significant.

```
linmodmap<-lm(cover~elev+tci+streamdist+beers, data=maple)
summary(linmodmap)
```

```
##
## Call:
## lm(formula = cover ~ elev + tci + streamdist + beers, data = maple)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.7869 -1.2983  0.3618  1.4014  5.2451
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  6.3218898  0.3604346  17.540  < 2e-16 ***
## elev        -0.0008868  0.0002606  -3.403 0.000703 ***
## tci         -0.0668631  0.0350647  -1.907 0.056939 .
## streamdist   0.0013256  0.0004696   2.823 0.004893 **
## beers       -0.3204370  0.1068951  -2.998 0.002814 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.988 on 718 degrees of freedom
## Multiple R-squared:  0.04174,    Adjusted R-squared:  0.0364
## F-statistic: 7.818 on 4 and 718 DF,  p-value: 3.603e-06
```

Create PDF in your applications with the Pdfcrowd HTML to PDF API PDFCROWD

All variables except tci were found to be significant (and tci value was very close to 0.05), so it can be said that elevation, stream distance, and transformed slope aspect affect red maple cover. This was not easy to see in the plots, which is why it's important to run actual numerical analyses. However, one of the assumptions to run a linear model is that the data is in fact linear. The R squared value for maples is very low, indicating the data is not linear. Looking at the plots for maples especially, it can be visually confirmed that the data is not linear and thus perhaps a lm isn't a suitable method of analysis. Variance in firs will be more accurately desribed because the data is more suited to the method of analysis.

2.

I re-examine the data using a GLM with a Poisson distribution, which makes cover a continuous variable instead of a discrete one.

```
firpoi<-glm(cover~elev+tci+streamdist+beers, data=fir, family='poisson')
summary(firpoi)
```

```
##
## Call:
## glm(formula = cover ~ elev + tci + streamdist + beers, family = "poisson",
##     data = fir)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.6202  -0.4589   0.1027   0.4045   1.9304
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept) -3.4450538  1.2541319  -2.747 0.006015 **
## elev         0.0026913  0.0007039   3.823 0.000132 ***
## tci          0.0680966  0.0484916   1.404 0.160231
## streamdist  -0.0002762  0.0003952  -0.699 0.484554
## beers        0.0147344  0.1324953   0.111 0.911452
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 41.274  on 43  degrees of freedom
## Residual deviance: 19.521  on 39  degrees of freedom
```

Create PDF in your applications with the Pdfcrowd HTML to PDF API PDFCROWD

```
## AIC: 186.7
##
## Number of Fisher Scoring iterations: 4
```

```
mappoi<-glm(cover~elev+tci+streamdist+beers, data=maple, family='poisson')
summary(mappoi)
```

```
##
## Call:
## glm(formula = cover ~ elev + tci + streamdist + beers, family = "poisson",
##     data = maple)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.4679  -0.6058   0.1613   0.5816   2.0898
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)  1.870e+00  8.102e-02  23.082  < 2e-16 ***
## elev        -1.719e-04  5.803e-05  -2.963  0.00305 **
## tci         -1.382e-02  8.140e-03  -1.698  0.08950 .
## streamdist   2.500e-04  1.019e-04   2.453  0.01419 *
## beers       -6.265e-02  2.376e-02  -2.637  0.00837 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 649.34  on 722  degrees of freedom
## Residual deviance: 625.28  on 718  degrees of freedom
## AIC: 3097.7
##
## Number of Fisher Scoring iterations: 4
```

Create PDF in your applications with the Pdfcrowd HTML to PDF API PDFCROWD

```r
fir_r2 = function(firpoi) {
                1 -  firpoi$deviance / firpoi$null.deviance
}
fir_r2(firpoi)
```

```
## [1] 0.527049
```

```r
map_r2 = function(mappoi) {
                1 -  mappoi$deviance / mappoi$null.deviance
}
map_r2(mappoi)
```

```
## [1] 0.03704802
```

Results from the glm are very similar to the results from the lm, with the same variables found to be significant for each tree type. The R squared value improves slightly for firs, while the value remains equally poor in the maples.

3.

In summary, the cover of fir trees seems to be influenced greatly by elevation. Seeing as we knew that firs were habitat specialists before the beginning of the exercise, this makes sense- we can infer they do well at high elevations. While many factors were identified as significantly affecting cover for maples, the models returned very low R squared values for these trees, suggesting the models weren't appropriate. Maples are generalists and seem to be found just about everywhere, and their plots showed minimal pattern, which would be expected of a generalist species that is capable of surviving in a wide array of conditions.

Create PDF in your applications with the Pdfcrowd HTML to PDF API PDFCROWD