# Homework 3

*Kevin Mack*

*2/3/2016*

## Univariate Assignment

1) Working with a tree dataset we are interested in how tree cover (local abundance measured as estimated horizontal cover (ie, relative area of shadow if sun is directly above) in classes from 1-10: 1=trace, 2=0-1%, 3=1-2%, 4=2-5%, 5=5-10%, 6=10-25%, 7=25-50%, 8=50-75%, 9=75-95%, 10=95-100%) of two species of trees is affected by a variety of other factors. The two tree species we are interested in are the Red maple, Acer rubrum, a habitat generalist, and the Frasier fir, Abies fraseri, a habitat specialist.

We will use univariate analysis to determine how cover (y) is influenced by elevation (elevation in meters from a digital elevation model , x1), water potential (topographic convergence index, or site "water potential"; measured as the upslope contributing area divided by the tangent of the slope angle, x2), stream distance ( distance of plot from the nearest permanent stream (meters), x3), disturbance (disturbance history (from a Park report); CORPLOG=corporate logging; SETTLE=concentrated settlement, VIRGIN="high in virgin attributes", LT-SEL=light or selective logging, x4), and "beers", which incidentaly, isnt as fun as it sounds ( transformed slope aspect ('heat load index'); 0 is SW (hottest), 2 is NE (coolest), x5)

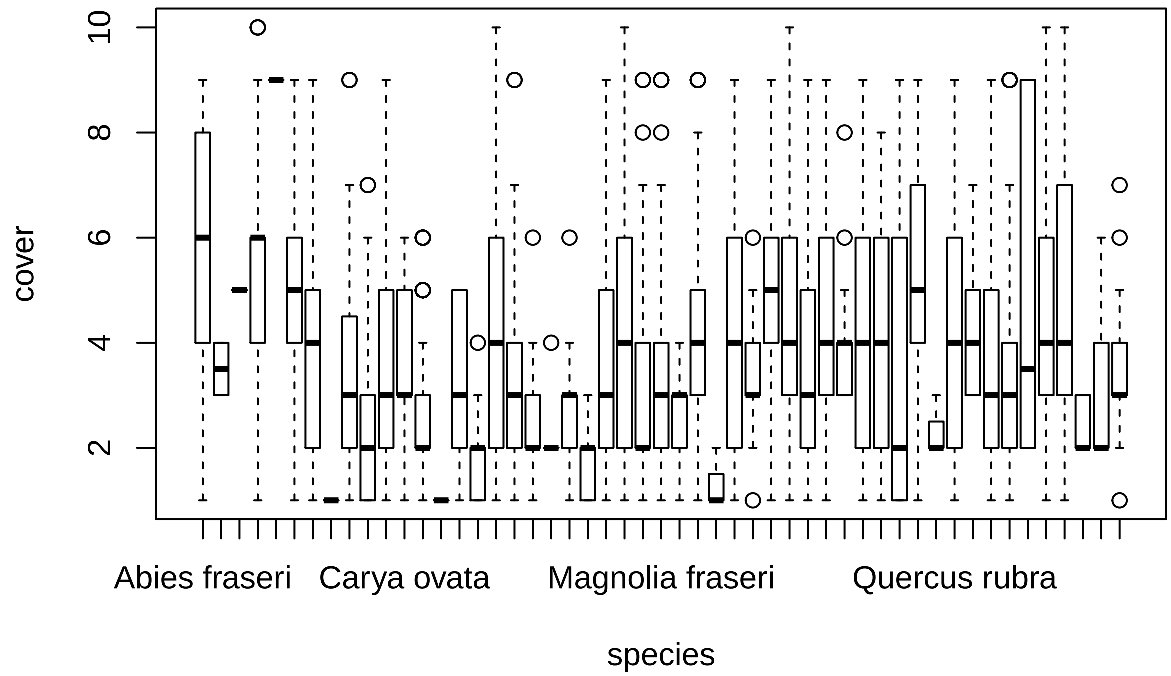
```
library(car)

trees = read.csv('https://raw.githubusercontent.com/dmcglinn/quant_methods/gh-pages/data/treedata_subse

names(trees)
```

```
## [1] "plotID"     "spcode"     "species"    "cover"      "elev"
## [6] "tci"        "streamdist" "disturb"    "beers"
```

```
sapply(trees, class)
```

```
##     plotID     spcode    species      cover       elev        tci
##   "factor"   "factor"   "factor"  "integer"  "numeric"  "numeric"
## streamdist    disturb      beers
##  "numeric"   "factor"  "numeric"
```

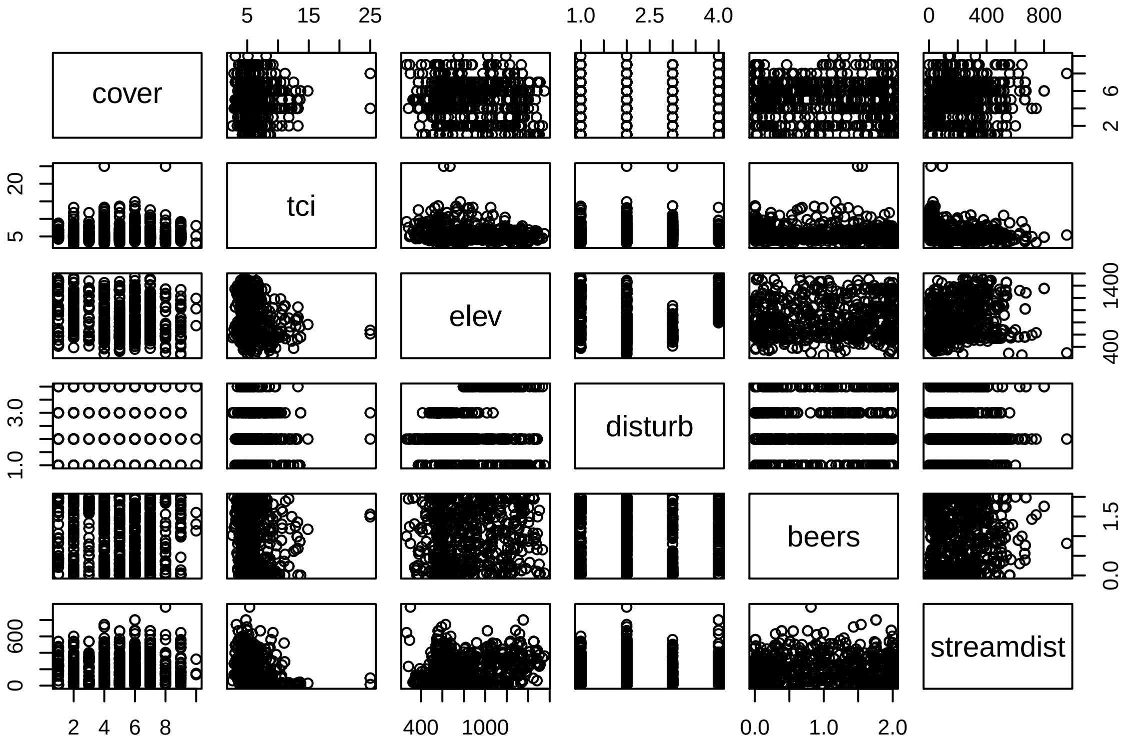
```
plot(cover~species, data = trees) #lots going on, doesnt tell us much
```



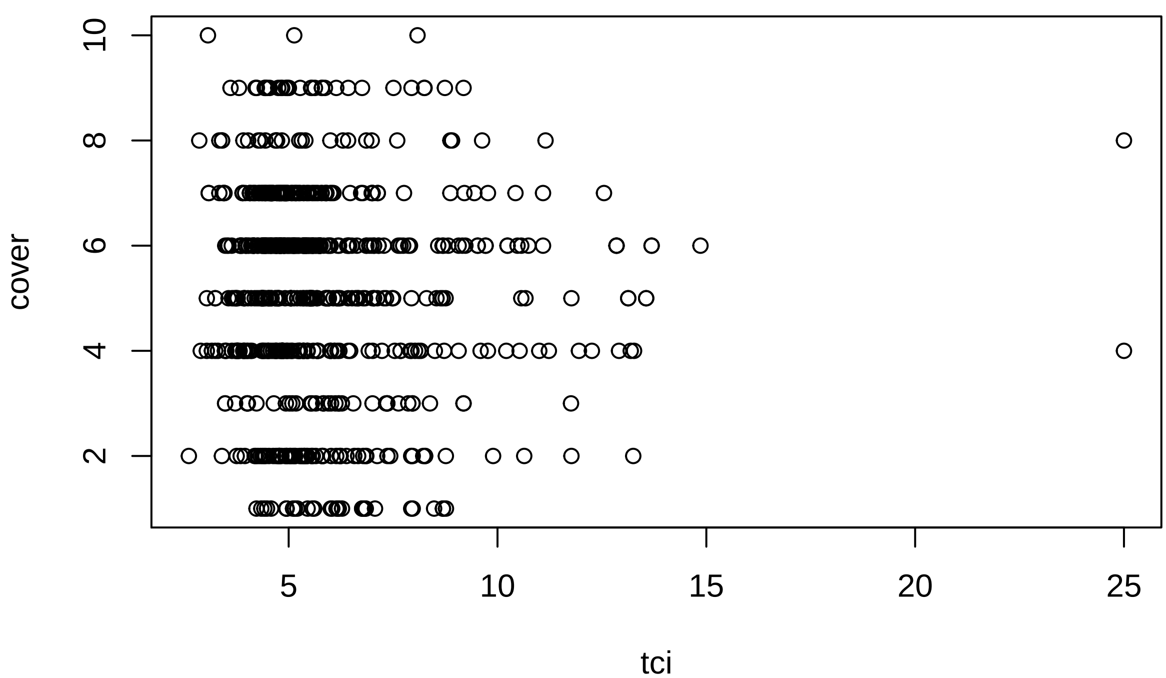

```
#subset data to look as species of interest
cols_of_interest = c('cover', 'tci', 'elev', 'disturb', 'beers', 'streamdist')
#pulls out variables we are interested in
acer = subset(trees, species == "Acer rubrum", select= cols_of_interest)
abies = subset(trees, species == "Abies fraseri", select=cols_of_interest)
```

Modeling Acer rubrum

```
plot(acer) #too much information, and only really interested in cover
```



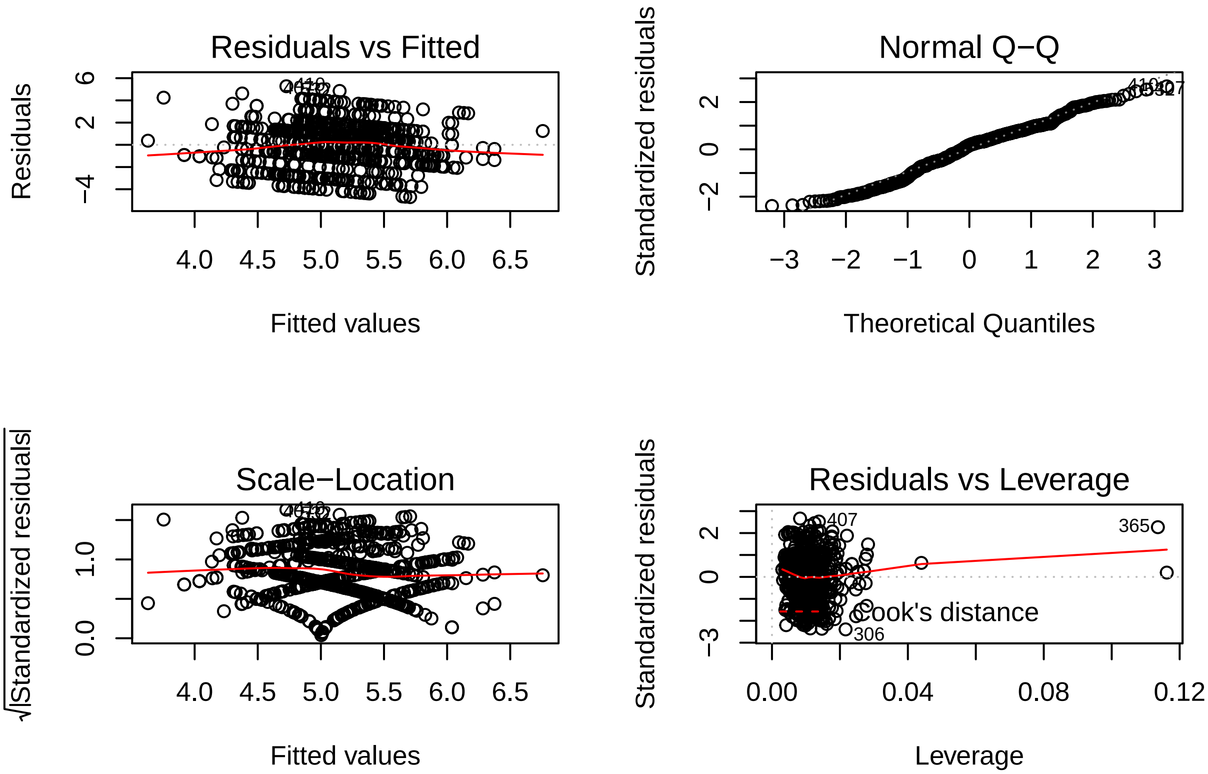

```r
par(mfrow=c(1,1))
plot(cover ~ tci + elev + disturb + beers + streamdist, data = acer) #better?
```




!!!!

```r
par(mfrow=c(1,1))

mod_acer = lm(cover ~ ., data=acer)
```

```r
par(mfrow=c(2,2))
plot(mod_acer)#plots of the model fit
```




```r
par(mfrow=c(1,1))
Anova(mod_acer, type=3)# elev, beers, and streamdist significantly affect cover
```

```
## Anova Table (Type III tests)
##
## Response: cover
##              Sum Sq  Df  F value    Pr(>F)
## (Intercept)  765.43   1 193.5096 < 2.2e-16 ***
## tci           12.58   1   3.1805  0.074947 .
## elev          40.44   1  10.2233  0.001448 **
## disturb        9.45   3   0.7962  0.496166
## beers         35.61   1   9.0034  0.002789 **
## streamdist    29.09   1   7.3531  0.006856 **
## Residuals   2828.21 715
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
summary(mod_acer) #4.805 ratio explained variation to unexplained, poor rsquared fit
```

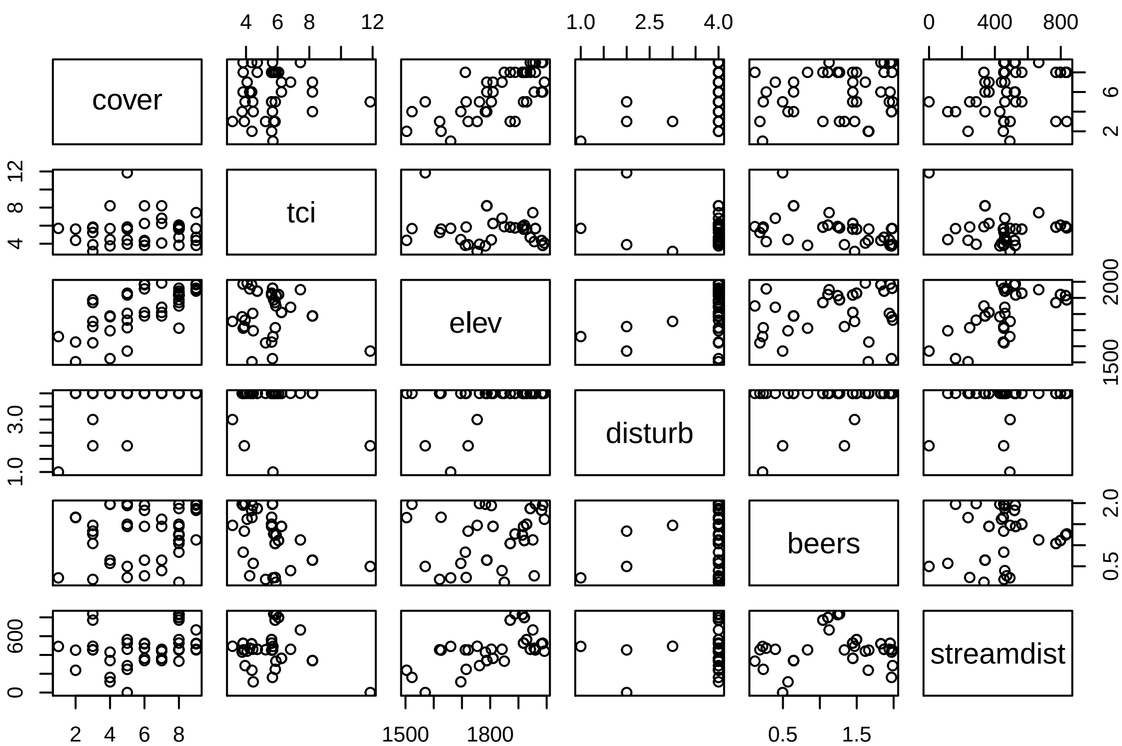
```
##
## Call:
## lm(formula = cover ~ ., data = acer)
##
```

```
## Residuals:
##     Min      1Q  Median      3Q     Max 
## -4.7073 -1.2446  0.3409  1.3575  5.2732 
## 
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)    
## (Intercept)    6.3502303  0.4564973  13.911  < 2e-16 ***
## tci           -0.0627613  0.0351922  -1.783  0.07495 .  
## elev          -0.0010108  0.0003161  -3.197  0.00145 ** 
## disturbLT-SEL  0.0829610  0.2166747   0.383  0.70192    
## disturbSETTLE -0.1044556  0.2804213  -0.372  0.70963    
## disturbVIRGIN  0.3088364  0.2518161   1.226  0.22044    
## beers         -0.3269597  0.1089662  -3.001  0.00279 ** 
## streamdist     0.0012895  0.0004756   2.712  0.00686 ** 
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 1.989 on 715 degrees of freedom
## Multiple R-squared:  0.04493,    Adjusted R-squared:  0.03558 
## F-statistic: 4.805 on 7 and 715 DF,  p-value: 2.669e-05
```

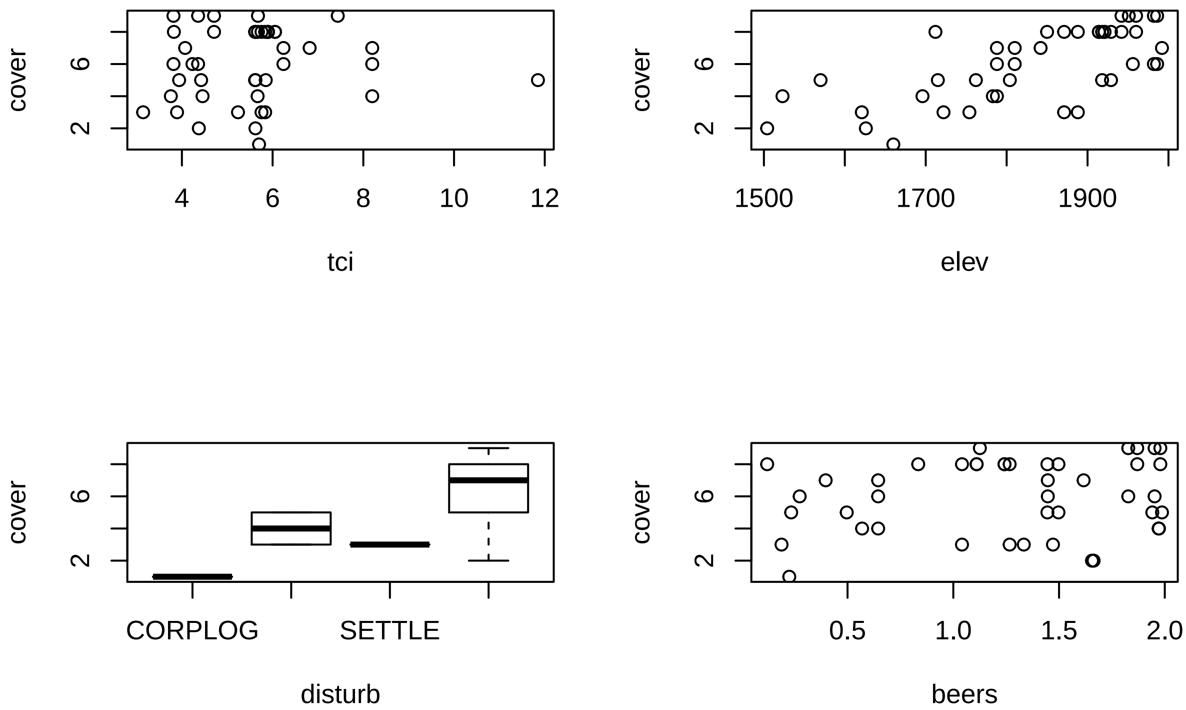
From this model, tree cover in Acer rubrum is significantly affected by elevation, heat load index, and distance from stream. Though, this model still has a lot of unexplained variance (f = 4.05) and a relitavely poor adjusted R-Squared (0.035). This model does not appear to explain cover in Acer rubrum very well, but because the species is a habitat generalist, we may have expected tree cover to influenced by a wide number of factors (and be difficult to model).

Modeling Abies fraseri

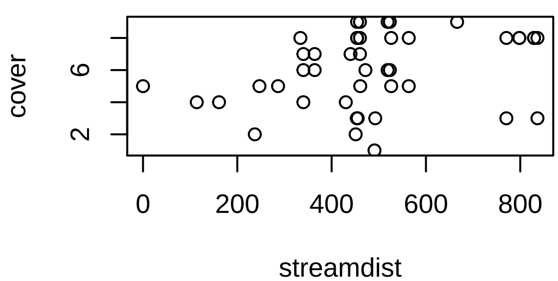
```r
plot(abies)
```



```r
par(mfrow=c(2,2))
plot(cover ~ tci + elev + disturb + beers + streamdist, data = abies)
```

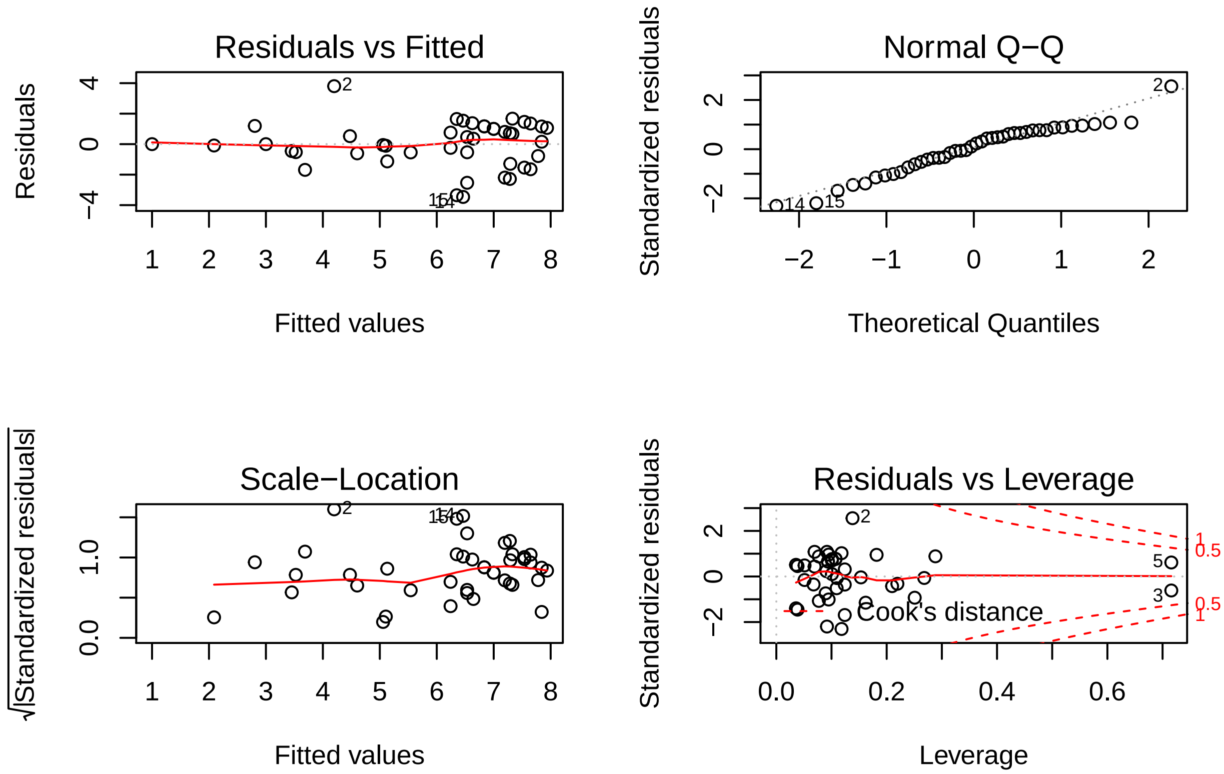



```r
par(mfrow=c(1,1))
```




```r
mod_abies = lm(cover ~ ., data=abies)
par(mfrow=c(2,2))
plot(mod_abies) #clear trend in elevation plot
```

```
## Warning: not plotting observations with leverage one:
##   1, 4

## Warning: not plotting observations with leverage one:
##   1, 4
```




```r
par(mfrow=c(1,1))
Anova(mod_abies, type=3) #only elevation significant
```

```
## Anova Table (Type III tests)
##
## Response: cover
##              Sum Sq Df F value    Pr(>F)
## (Intercept) 59.401  1 23.1710 2.652e-05 ***
## tci          5.667  1  2.2105    0.1458
## elev        61.618  1 24.0358 2.022e-05 ***
## disturb     10.089  3  1.3118    0.2855
## beers        0.014  1  0.0056    0.9406
## streamdist   1.636  1  0.6382    0.4296
## Residuals   92.289 36
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
summary(mod_abies) # 7.17 ratio explained to unexplained variation, 0.5 rsquared
```

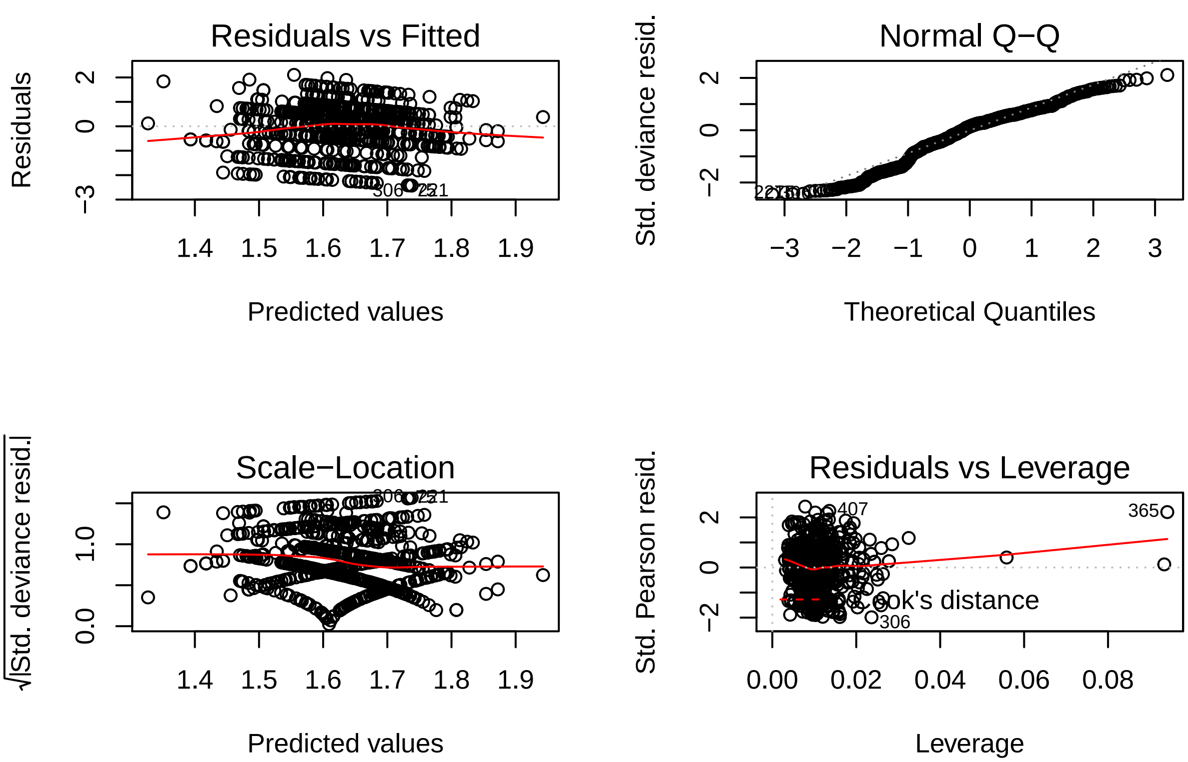
```
##
## Call:
## lm(formula = cover ~ ., data = abies)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4630 -0.6472  0.0788  1.0872  3.8017
##
```

```
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -20.561173   4.271449  -4.814 2.65e-05 ***
## tci             0.287641   0.193467   1.487   0.1458
## elev            0.012370   0.002523   4.903 2.02e-05 ***
## disturbLT-SEL   2.188367   2.097905   1.043   0.3038
## disturbSETTLE   1.527604   2.341471   0.652   0.5183
## disturbVIRGIN   3.025596   1.735921   1.743   0.0899 .
## beers           0.037551   0.500269   0.075   0.9406
## streamdist     -0.001266   0.001585  -0.799   0.4296
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.601 on 36 degrees of freedom
## Multiple R-squared:  0.5824, Adjusted R-squared:  0.5011
## F-statistic: 7.171 on 7 and 36 DF,  p-value: 2.215e-05
```

From this model, tree cover in Abies fraseri is significantly affected only by elevation. This model better explains variation (f = 7.17) and has a better fit (R-Squared = 0.5) than the model of Acer rubrum. Because Abies fraseri is a habitat specalist, with a seemingly higher fitness in higher elevations, it is reasonable that changes in elevation are more significant than other factors.

2) Re-Examine with Poisson Distribution

```r
acer_glm = glm(cover ~ . , data= acer, family='poisson')
par(mfrow=c(2,2))
plot(acer_glm)
```



```r
par(mfrow=c(1,1))

rsq=function(glm_mod){
  1-glm_mod$deviance/glm_mod$null.deviance
}

acer_glm #new model
```

```
##
## Call:  glm(formula = cover ~ ., family = "poisson", data = acer)
##
## Coefficients:
##   (Intercept)            tci           elev  disturbLT-SEL  disturbSETTLE
##     1.8730109     -0.0129660     -0.0001961      0.0183973     -0.0173856
## disturbVIRGIN          beers      streamdist
##     0.0631125     -0.0639106      0.0002428
##
## Degrees of Freedom: 722 Total (i.e. Null);  715 Residual
## Null Deviance:       649.3
## Residual Deviance: 623.4     AIC: 3102
```

```r
summary(acer_glm) #same significant factors
```

```
##
## Call:
## glm(formula = cover ~ ., family = "poisson", data = acer)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.4282  -0.5903   0.1391   0.5786   2.1038
##
## Coefficients:
##                 Estimate Std. Error z value Pr(>|z|)
## (Intercept)    1.873e+00  1.023e-01  18.315  < 2e-16 ***
## tci           -1.297e-02  8.159e-03  -1.589  0.11202
## elev          -1.961e-04  7.047e-05  -2.783  0.00538 **
## disturbLT-SEL  1.840e-02  4.880e-02   0.377  0.70619
## disturbSETTLE -1.739e-02  6.253e-02  -0.278  0.78099
## disturbVIRGIN  6.311e-02  5.638e-02   1.119  0.26293
## beers         -6.391e-02  2.423e-02  -2.638  0.00834 **
## streamdist     2.428e-04  1.030e-04   2.357  0.01843 *
## ---
## 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: 623.38  on 715  degrees of freedom
## AIC: 3101.8
##
## Number of Fisher Scoring iterations: 4
```

```r
rsq(acer_glm) #slightly better rsquared (0.04), still a poorly fitted model
```

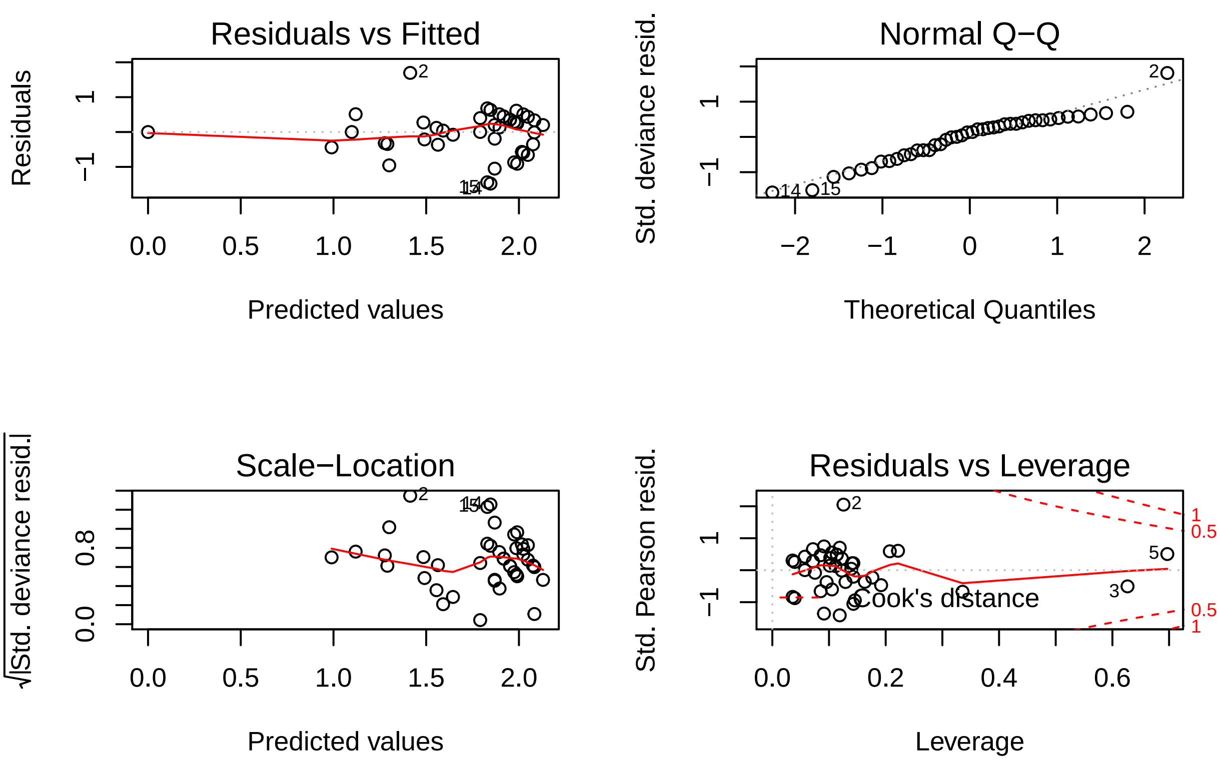
```
## [1] 0.03997917
```

```r
anova(mod_acer, acer_glm) # much lower RSS in glm indicates better fit
```

```
## Analysis of Variance Table
##
## Model 1: cover ~ tci + elev + disturb + beers + streamdist
## Model 2: cover ~ tci + elev + disturb + beers + streamdist
##   Res.Df     RSS Df Sum of Sq F Pr(>F)
## 1    715 2828.21
## 2    715  623.38  0    2204.8
```

```r
abies_glm = glm(cover ~ . , data= abies, family='poisson')
par(mfrow=c(2,2))
plot(abies_glm) #them be some squiggly lines
```

```
## Warning: not plotting observations with leverage one:
##   1, 4

## Warning: not plotting observations with leverage one:
##   1, 4
```



```r
par(mfrow=c(1,1))


abies_glm
```

```
##
## Call:  glm(formula = cover ~ ., family = "poisson", data = abies)
##
## Coefficients:
##   (Intercept)            tci           elev  disturbLT-SEL  disturbSETTLE
##    -4.1157009      0.0568868      0.0023508      1.2440008      1.0440232
## disturbVIRGIN          beers     streamdist
##     1.4002993     -0.0165548     -0.0002186
##
## Degrees of Freedom: 43 Total (i.e. Null);  36 Residual
## Null Deviance:       41.27
## Residual Deviance: 16.13     AIC: 189.3
```

```r
summary(abies_glm) #same significant variable
```

```
##
## Call:
## glm(formula = cover ~ ., family = "poisson", data = abies)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -1.47931  -0.35524   0.08027   0.36453   1.69535
##
## Coefficients:
##                 Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -4.1157009  1.5505526  -2.654  0.00795 **
## tci            0.0568868  0.0524222   1.085  0.27785
## elev           0.0023508  0.0007292   3.224  0.00126 **
## disturbLT-SEL  1.2440008  1.0827736   1.149  0.25060
## disturbSETTLE  1.0440232  1.1644892   0.897  0.36996
## disturbVIRGIN  1.4002993  1.0171140   1.377  0.16859
## beers         -0.0165548  0.1326724  -0.125  0.90070
## streamdist    -0.0002186  0.0003969  -0.551  0.58176
## ---
## 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: 16.126  on 36  degrees of freedom
## AIC: 189.3
##
## Number of Fisher Scoring iterations: 4
```

```r
rsq(abies_glm) #better rsquared (0.60)
```

```
## [1] 0.60931
```

```r
anova(mod_abies, abies_glm)#again, lowered residual sum of squares in glm indicates better fit
```

```
## Analysis of Variance Table
##
## Model 1: cover ~ tci + elev + disturb + beers + streamdist
## Model 2: cover ~ tci + elev + disturb + beers + streamdist
##   Res.Df    RSS Df Sum of Sq F Pr(>F)
## 1     36 92.289
## 2     36 16.126  0    76.164
```

In both species, the glm model has improved the rsquared fit. Addationally, the glm model had lower residual sum of squares compared to the olm model.

3) The biological story told by these data paints a picture of two different habitat selection modes. The habitat specailist, Abies fraseri fits a model of tree coverage that is best explained by elevation out of all of the other factors. As elevation increases, so does cover, and as a proxy the fitness of this species. Other factors such as stream distance or heat load are less significant to this model, likely because this species has specialized to living in higher elevations. The alternate story is that of the habitat generalist, Acer rubrum. In this species, cover can be explained most effectively by elevation, but also by heat load index, and distance from the nearest stream. The fitness of a species that does not fulfill any specific niche is likely to be dependent on a variety of factors.