Univariate Assignment

Nick Weber

1/24/2018

```
#1. Explanatory analysis for Acer rubrum and Abies fraseri...


library(car)
trees <- read.csv("treedata_subset.csv", header=TRUE)

#ACER RUBRUM

a <- trees$elev[trees$species=="Acer rubrum"]
b <- trees$tci[trees$species=="Acer rubrum"]
c <- trees$streamdist[trees$species=="Acer rubrum"]
d <- trees$disturb[trees$species=="Acer rubrum"]
e <- trees$beers[trees$species=="Acer rubrum"]

acer_mod1 <- lm(cover[species=="Acer rubrum"] ~ a + b + c + as.factor(d) + e, data=tree
s)
summary(acer_mod1) #Adj.R2=0.036
```
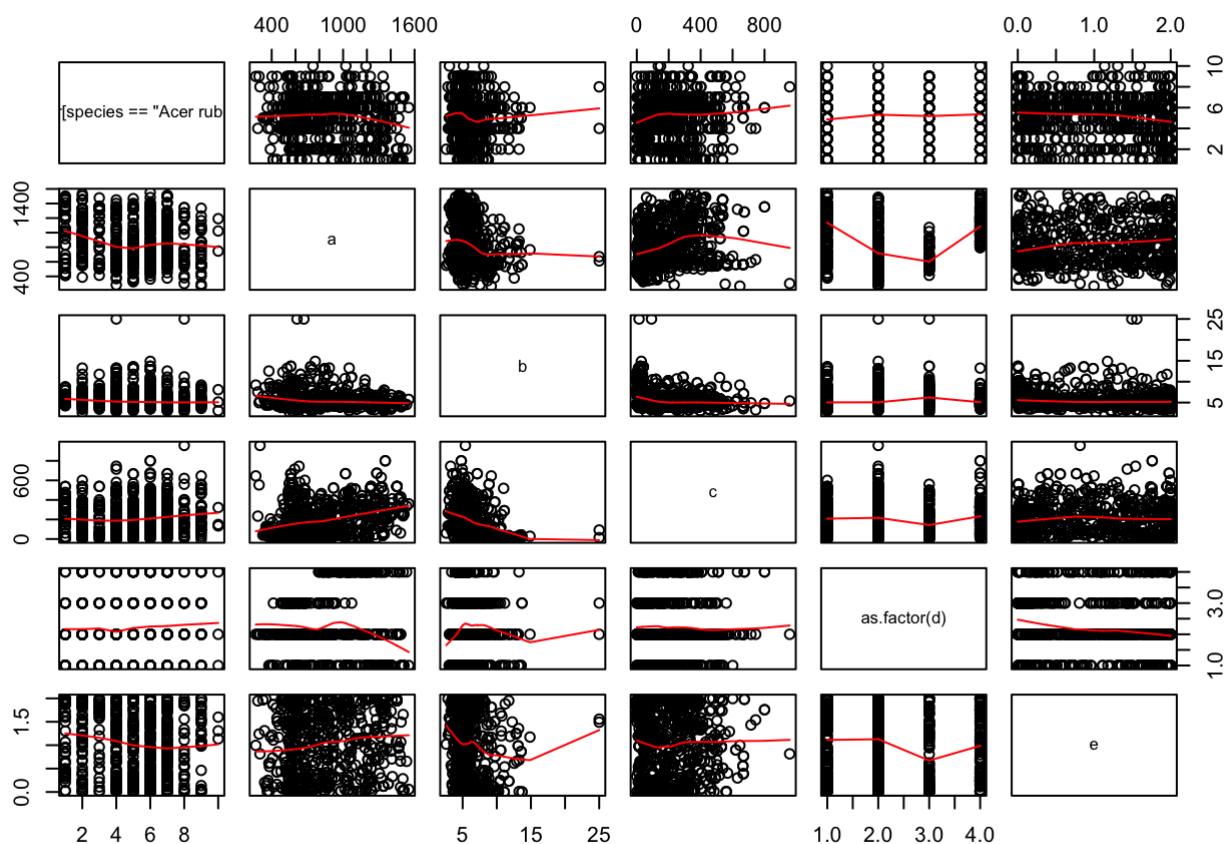
```
##
## Call:
## lm(formula = cover[species == "Acer rubrum"] ~ a + b + c + as.factor(d) +
##     e, data = trees)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.7073 -1.2446  0.3409  1.3575  5.2732
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)        6.3502303  0.4564973  13.911  < 2e-16 ***
## a                 -0.0010108  0.0003161  -3.197  0.00145 **
## b                 -0.0627613  0.0351922  -1.783  0.07495 .
## c                  0.0012895  0.0004756   2.712  0.00686 **
## as.factor(d)LT-SEL  0.0829610  0.2166747   0.383  0.70192
## as.factor(d)SETTLE -0.1044556  0.2804213  -0.372  0.70963
## as.factor(d)VIRGIN  0.3088364  0.2518161   1.226  0.22044
## e                 -0.3269597  0.1089662  -3.001  0.00279 **
## ---
## 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
```

```
Anova(acer_mod1, type=3)
```

```
## Anova Table (Type III tests)
##
## Response: cover[species == "Acer rubrum"]
##                 Sum Sq  Df  F value      Pr(>F)
## (Intercept)     765.43   1 193.5096 < 2.2e-16 ***
## a                40.44   1  10.2233  0.001448 **
## b                12.58   1   3.1805  0.074947 .
## c                29.09   1   7.3531  0.006856 **
## as.factor(d)      9.45   3   0.7962  0.496166
## e                35.61   1   9.0034  0.002789 **
## Residuals      2828.21 715
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#The p-values generated by Anova() are the same as those generated by summary(). Howeve
r, summary() generates p-values for all of the categories in "disturb", while Anova() gr
oups those categories into one p-value.

pairs(cover[species=="Acer rubrum"] ~ a + b + c + as.factor(d) + e, data=trees, panel=pa
nel.smooth)
```

```
#Removing least sig term first (d <- disturb)...

acer_mod2 <- lm(cover[species=="Acer rubrum"] ~ a + b + c + e, data=trees)
summary(acer_mod2) #Adj.R2=0.036
```

```
##
## Call:
## lm(formula = cover[species == "Acer rubrum"] ~ a + b + c + e,
##     data = trees)
##
## 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 ***
## a           -0.0008868  0.0002606  -3.403 0.000703 ***
## b           -0.0668631  0.0350647  -1.907 0.056939 .
## c            0.0013256  0.0004696   2.823 0.004893 **
## e           -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
```

```
Anova(acer_mod2, type=3)
```

```
## Anova Table (Type III tests)
##
## Response: cover[species == "Acer rubrum"]
##               Sum Sq  Df  F value    Pr(>F)
## (Intercept) 1215.84    1 307.6386 < 2.2e-16 ***
## a             45.78    1  11.5830 0.0007025 ***
## b             14.37    1   3.6361 0.0569390 .
## c             31.49    1   7.9681 0.0048925 **
## e             35.51    1   8.9861 0.0028143 **
## Residuals   2837.66  718
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#Again removing least sig term (b <- tci)...

acer_mod3 <- lm(cover[species=="Acer rubrum"] ~ a + c + e, data=trees)
summary(acer_mod3) #All terms (a, c, e) are significant, but Adj.R2=0.033
```

```
##
## Call:
## lm(formula = cover[species == "Acer rubrum"] ~ a + c + e, data = trees)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.9339 -1.3307  0.2865  1.3691  5.2839
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  5.8119219  0.2420805  24.008  < 2e-16 ***
## a           -0.0008226  0.0002589  -3.178 0.001548 **
## c            0.0015415  0.0004566   3.376 0.000774 ***
## e           -0.3034710  0.1067192  -2.844 0.004587 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.992 on 719 degrees of freedom
## Multiple R-squared:  0.03688,    Adjusted R-squared:  0.03287
## F-statistic: 9.178 on 3 and 719 DF,  p-value: 5.775e-06
```

```
Anova(acer_mod3, type=3)
```

```
## Anova Table (Type III tests)
##
## Response: cover[species == "Acer rubrum"]
##              Sum Sq  Df  F value     Pr(>F)
## (Intercept) 2286.36   1 576.3946 < 2.2e-16 ***
## a             40.05   1  10.0977  0.001548 **
## c             45.22   1  11.4000  0.000774 ***
## e             32.08   1   8.0863  0.004587 **
## Residuals   2852.03 719
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#Let's look at that same model, but accounting for potential interactions...

acer_modinteract <- lm(cover[species=="Acer rubrum"] ~ a + c + e + a*e, data=trees)
summary(acer_modinteract) #All terms (a, c, e, a*e) are significant, and Adj.R2=0.061
```
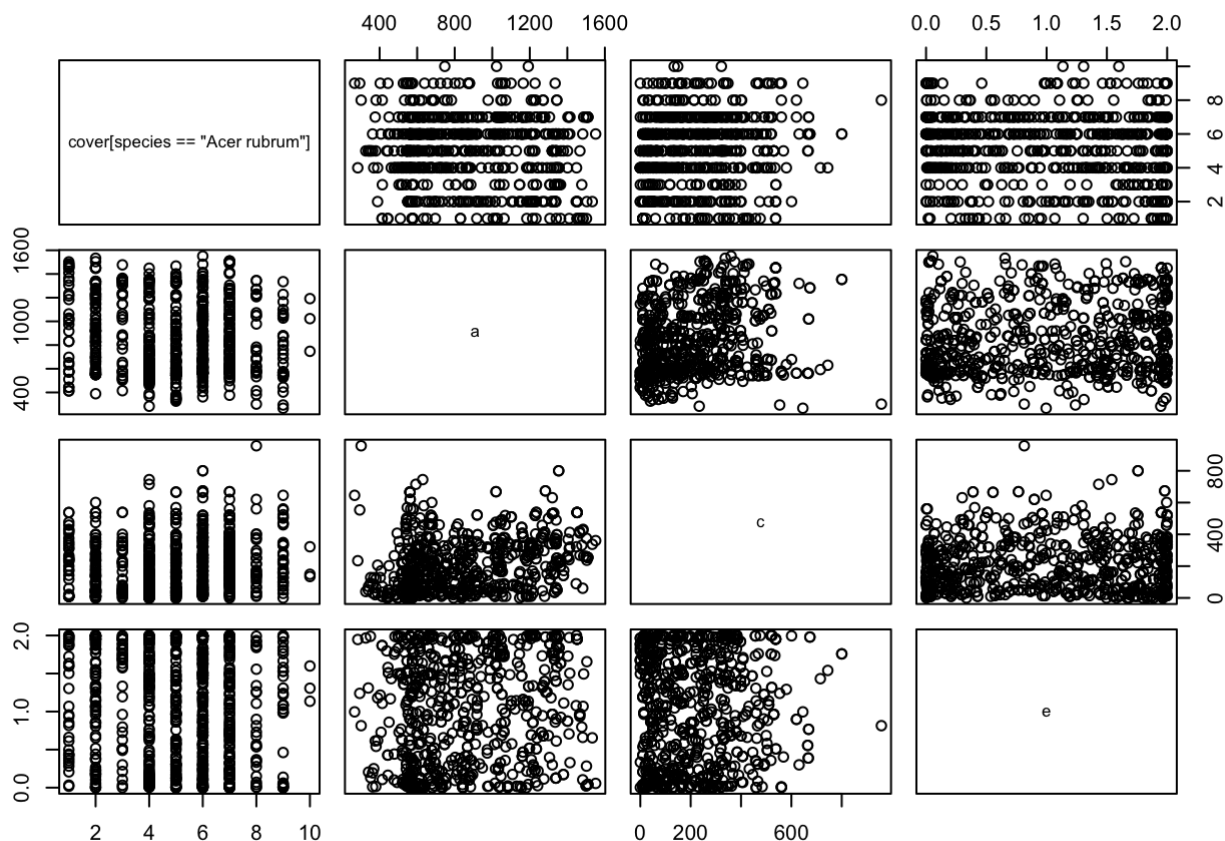
```
##
## Call:
## lm(formula = cover[species == "Acer rubrum"] ~ a + c + e + a *
##     e, data = trees)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.9675 -1.2255  0.2472  1.3318  5.3015
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.3684250  0.3884493  11.246  < 2e-16 ***
## a            0.0008955  0.0004452   2.011 0.044651 *
## c            0.0017225  0.0004516   3.814 0.000149 ***
## e            1.0924924  0.3145520   3.473 0.000545 ***
## a:e         -0.0016507  0.0003505  -4.709 2.99e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.963 on 718 degrees of freedom
## Multiple R-squared:  0.06574,    Adjusted R-squared:  0.06053
## F-statistic: 12.63 on 4 and 718 DF,  p-value: 6.155e-10
```

```
Anova(acer_modinteract, type=3)
```

```
## Anova Table (Type III tests)
##
## Response: cover[species == "Acer rubrum"]
##             Sum Sq  Df  F value     Pr(>F)
## (Intercept)  487.31   1 126.4681 < 2.2e-16 ***
## a             15.59   1   4.0459 0.0446509 *
## c             56.05   1  14.5464 0.0001485 ***
## e             46.48   1  12.0629 0.0005452 ***
## a:e           85.44   1  22.1748 2.988e-06 ***
## Residuals  2766.59 718
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
pairs(cover[species=="Acer rubrum"] ~ a + c + e + a*e, data=trees)
```

```
#Now attempting just single variable models, with the 3 variables that were significant
 in the multiple regression.

acer_moda <- lm(cover[species=="Acer rubrum"] ~ a, data=trees)
summary(acer_moda) #Adj.R2=0.01
```

```
##
## Call:
## lm(formula = cover[species == "Acer rubrum"] ~ a, data = trees)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.4433 -1.3237  0.5212  1.2521  5.1009
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  5.7319818  0.2275539  25.190  < 2e-16 ***
## a           -0.0006987  0.0002505  -2.789  0.00543 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.016 on 721 degrees of freedom
## Multiple R-squared:  0.01067,    Adjusted R-squared:  0.009301
## F-statistic: 7.778 on 1 and 721 DF,  p-value: 0.005428
```

```
acer_modc <- lm(cover[species=="Acer rubrum"] ~ c, data=trees)
summary(acer_modc) #Adj.R2=0.01
```

```
##
## Call:
## lm(formula = cover[species == "Acer rubrum"] ~ c, data = trees)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.4895 -1.1789  0.2168  1.1007  4.9614
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 4.8881261  0.1231777  39.684   <2e-16 ***
## c           0.0011187  0.0004466   2.505   0.0125 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.018 on 721 degrees of freedom
## Multiple R-squared:  0.008626,   Adjusted R-squared:  0.007251
## F-statistic: 6.273 on 1 and 721 DF,  p-value: 0.01248
```

```
acer_mode <- lm(cover[species=="Acer rubrum"] ~ e, data=trees)
summary(acer_mode) #Adj.R2=0.01
```

```
##
## Call:
## lm(formula = cover[species == "Acer rubrum"] ~ e, data = trees)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.4760 -1.3192  0.5177  1.1995  5.0634
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   5.4823     0.1325  41.380  < 2e-16 ***
## e            -0.3413     0.1067  -3.197  0.00145 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.012 on 721 degrees of freedom
## Multiple R-squared:  0.01398,    Adjusted R-squared:  0.01261
## F-statistic: 10.22 on 1 and 721 DF,  p-value: 0.001447
```

```
#For Acer rubrum, the best fitting model (acer_modinteract) contains the explanatory var
iables elev, streamdist, and beers, with an interaction between elev and beers. While th
ese variables are the most important, the model does not do a very good job of explainin
g cover (Adj.R2=0.061).

#Checking to see if any assumptions are violated for Acer rubrum...
par(mfrow=c(2,2))
plot(acer_modinteract)
```
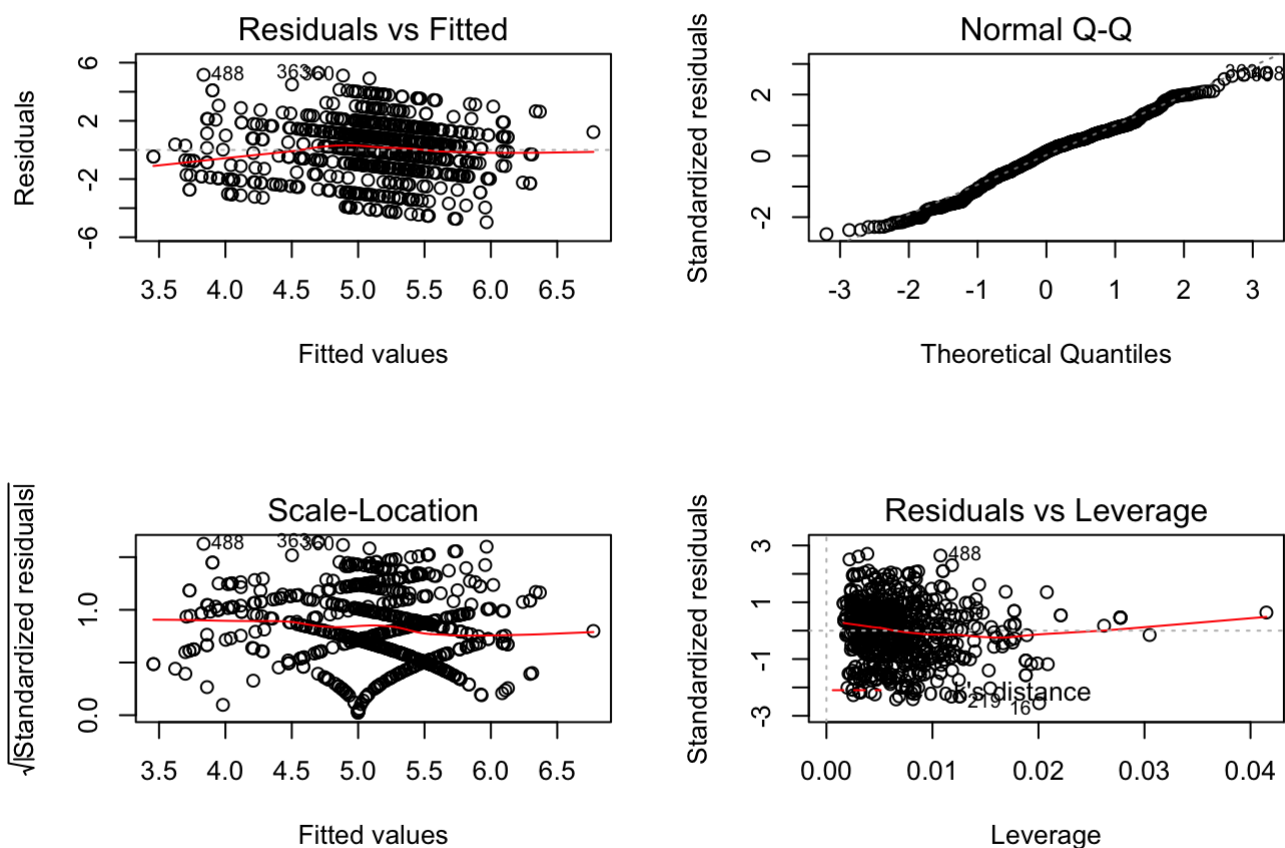


```
#It appears that no assumptions are violated.
```

```
#ABIES FRASERI

a.fras <- trees$elev[trees$species=="Abies fraseri"]
b.fras <- trees$tci[trees$species=="Abies fraseri"]
c.fras <- trees$streamdist[trees$species=="Abies fraseri"]
d.fras <- trees$disturb[trees$species=="Abies fraseri"]
e.fras <- trees$beers[trees$species=="Abies fraseri"]

abies_mod1 <- lm(cover[species=="Abies fraseri"] ~ a.fras + b.fras + c.fras + as.factor
(d.fras) + e.fras, data=trees)
summary(abies_mod1) #Adj.R2=0.50
```
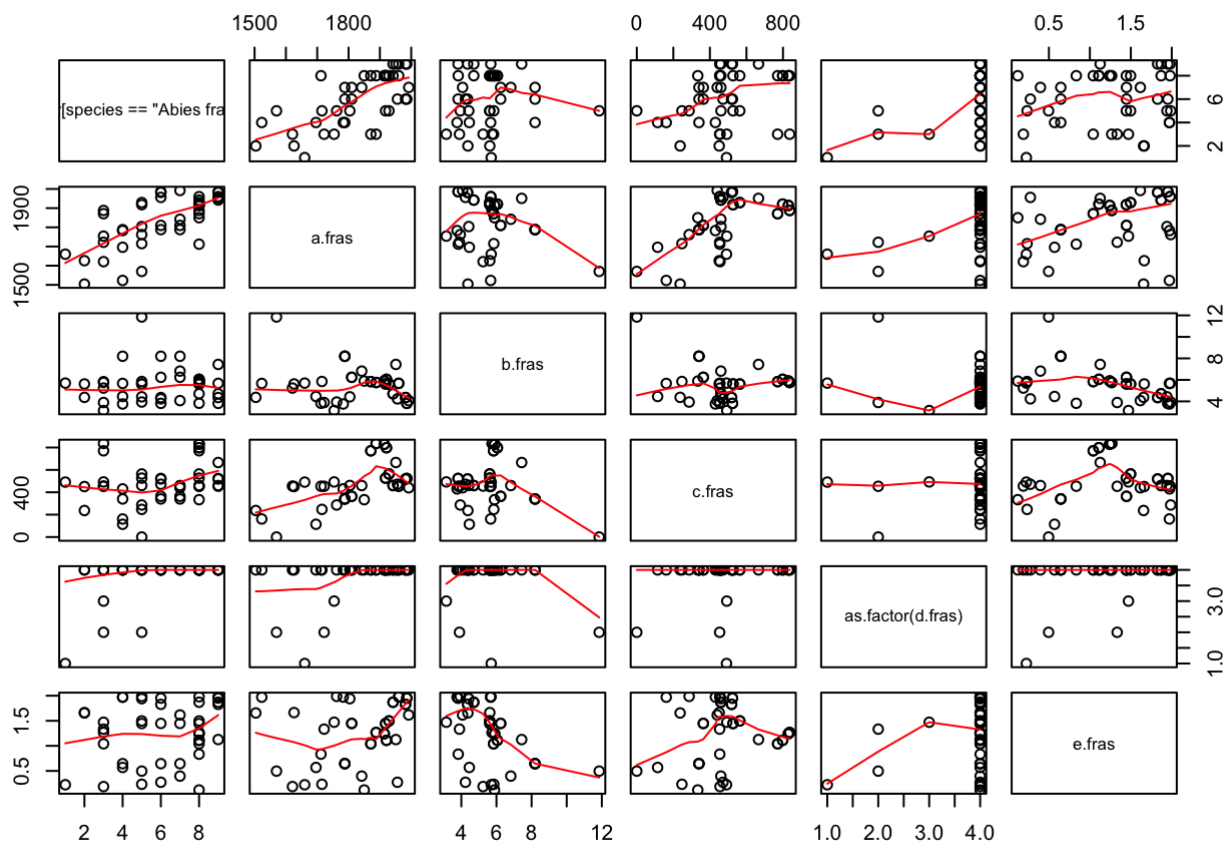
```
##
## Call:
## lm(formula = cover[species == "Abies fraseri"] ~ a.fras + b.fras +
##     c.fras + as.factor(d.fras) + e.fras, data = trees)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4630 -0.6472  0.0788  1.0872  3.8017
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)            -20.561173   4.271449  -4.814 2.65e-05 ***
## a.fras                   0.012370   0.002523   4.903 2.02e-05 ***
## b.fras                   0.287641   0.193467   1.487   0.1458
## c.fras                  -0.001266   0.001585  -0.799   0.4296
## as.factor(d.fras)LT-SEL  2.188367   2.097905   1.043   0.3038
## as.factor(d.fras)SETTLE  1.527604   2.341471   0.652   0.5183
## as.factor(d.fras)VIRGIN  3.025596   1.735921   1.743   0.0899 .
## e.fras                   0.037551   0.500269   0.075   0.9406
## ---
## 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
```

```
Anova(abies_mod1, type=3)
```

```
## Anova Table (Type III tests)
##
## Response: cover[species == "Abies fraseri"]
##                   Sum Sq Df F value    Pr(>F)
## (Intercept)       59.401  1 23.1710 2.652e-05 ***
## a.fras            61.618  1 24.0358 2.022e-05 ***
## b.fras             5.667  1  2.2105    0.1458
## c.fras             1.636  1  0.6382    0.4296
## as.factor(d.fras) 10.089  3  1.3118    0.2855
## e.fras             0.014  1  0.0056    0.9406
## Residuals         92.289 36
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
pairs(cover[species=="Abies fraseri"] ~ a.fras + b.fras + c.fras + as.factor(d.fras) +
  e.fras, data=trees, panel=panel.smooth)
```

```
#Removing least sig term first (e <- beers)...

abies_mod2 <- lm(cover[species=="Abies fraseri"] ~ a.fras + b.fras + c.fras + as.factor
(d.fras), data=trees)
summary(abies_mod2) #Adj.R2=0.51
```

```
##
## Call:
## lm(formula = cover[species == "Abies fraseri"] ~ a.fras + b.fras +
##     c.fras + as.factor(d.fras), data = trees)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4611 -0.6713  0.0899  1.0922  3.7785
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)             -20.575773   4.209291  -4.888 1.99e-05 ***
## a.fras                    0.012409   0.002437   5.091 1.06e-05 ***
## b.fras                    0.281194   0.171010   1.644   0.1086
## c.fras                   -0.001275   0.001560  -0.817   0.4192
## as.factor(d.fras)LT-SEL   2.226671   2.007369   1.109   0.2745
## as.factor(d.fras)SETTLE   1.554346   2.282898   0.681   0.5002
## as.factor(d.fras)VIRGIN   3.057250   1.661139   1.840   0.0737 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.579 on 37 degrees of freedom
## Multiple R-squared:  0.5823, Adjusted R-squared:  0.5146
## F-statistic: 8.596 on 6 and 37 DF,  p-value: 7.064e-06
```

```
Anova(abies_mod2, type=3)
```

```
## Anova Table (Type III tests)
##
## Response: cover[species == "Abies fraseri"]
##                   Sum Sq Df F value     Pr(>F)
## (Intercept)       59.609  1 23.8943 1.992e-05 ***
## a.fras            64.662  1 25.9199 1.064e-05 ***
## b.fras             6.745  1  2.7038    0.1086
## c.fras             1.665  1  0.6674    0.4192
## as.factor(d.fras) 10.679  3  1.4269    0.2504
## Residuals         92.304 37
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#Again removing least sig term (c <- streamdist)...

abies_mod3 <- lm(cover[species=="Abies fraseri"] ~ a.fras + b.fras + as.factor(d.fras),
 data=trees)
summary(abies_mod3) #Adj.R2=0.52
```

```
## 
## Call:
## lm(formula = cover[species == "Abies fraseri"] ~ a.fras + b.fras + 
##     as.factor(d.fras), data = trees)
## 
## Residuals:
##     Min      1Q  Median      3Q     Max 
## -3.8466 -0.6996  0.2255  0.8636  3.6846 
## 
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)    
## (Intercept)             -19.33664    3.90928  -4.946 1.57e-05 ***
## a.fras                    0.01128    0.00200   5.641 1.77e-06 ***
## b.fras                    0.28248    0.17025   1.659   0.1053    
## as.factor(d.fras)LT-SEL   2.54484    1.96059   1.298   0.2021    
## as.factor(d.fras)SETTLE   1.66160    2.26912   0.732   0.4685    
## as.factor(d.fras)VIRGIN   3.25929    1.63542   1.993   0.0535 .  
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 1.573 on 38 degrees of freedom
## Multiple R-squared:  0.5748, Adjusted R-squared:  0.5188 
## F-statistic: 10.27 on 5 and 38 DF,  p-value: 2.774e-06
```

```
Anova(abies_mod3, type=3)
```

```
## Anova Table (Type III tests)
## 
## Response: cover[species == "Abies fraseri"]
##                   Sum Sq Df F value    Pr(>F)    
## (Intercept)       60.502  1 24.4663 1.57e-05 ***
## a.fras            78.687  1 31.8202 1.77e-06 ***
## b.fras             6.807  1  2.7528    0.1053    
## as.factor(d.fras) 12.021  3  1.6204    0.2007    
## Residuals         93.969 38                      
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```
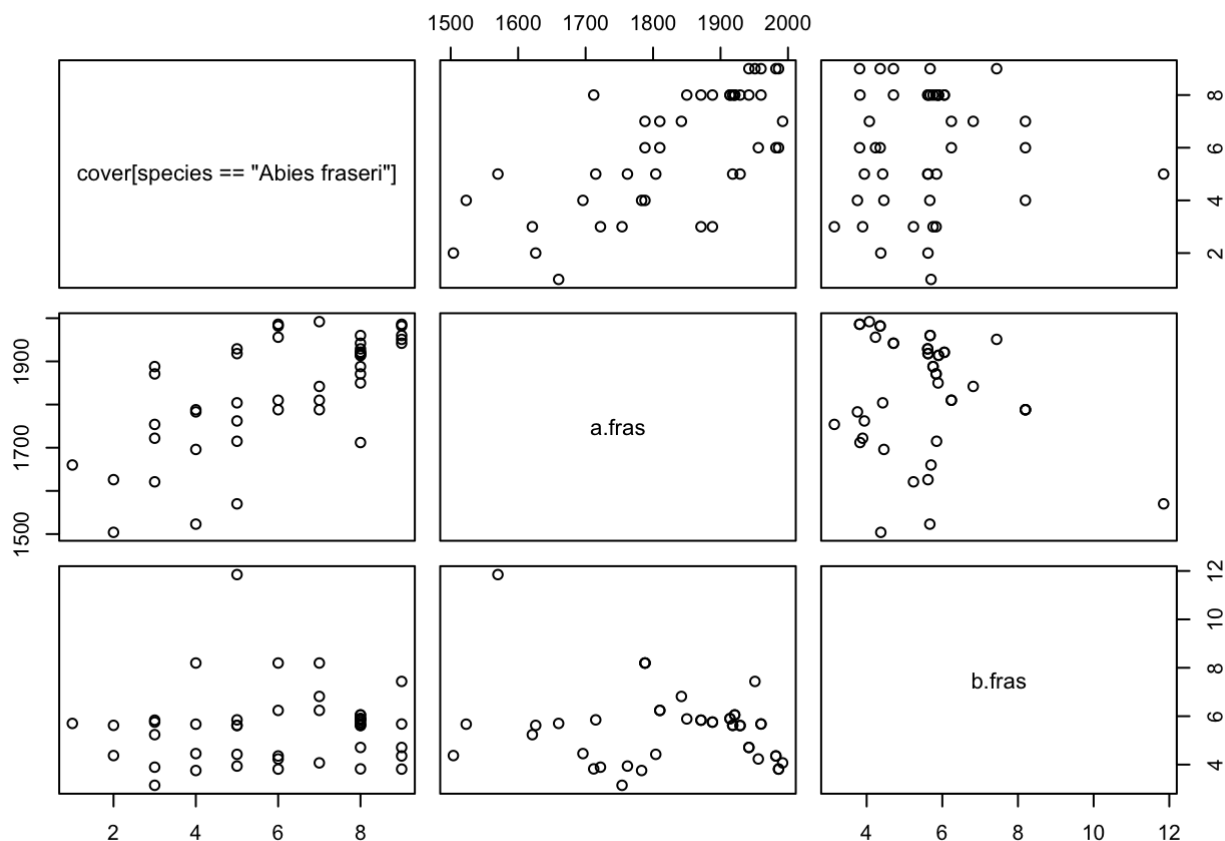
```
#Again removing least sig term (d <- disturb)...

abies_mod4 <- lm(cover[species=="Abies fraseri"] ~ a.fras + b.fras, data=trees)
summary(abies_mod4) #Adj.R2=0.50
```

```
##
## Call:
## lm(formula = cover[species == "Abies fraseri"] ~ a.fras + b.fras,
##     data = trees)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.7819 -1.1346  0.3731  0.8880  4.0268
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -18.789840   3.767380  -4.988 1.17e-05 ***
## a.fras        0.012616   0.001899   6.642 5.29e-08 ***
## b.fras        0.304539   0.161094   1.890   0.0658 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.608 on 41 degrees of freedom
## Multiple R-squared:  0.5204, Adjusted R-squared:  0.497
## F-statistic: 22.24 on 2 and 41 DF,  p-value: 2.876e-07
```

```
Anova(abies_mod4, type=3)
```

```
## Anova Table (Type III tests)
##
## Response: cover[species == "Abies fraseri"]
##             Sum Sq Df F value     Pr(>F)
## (Intercept)  64.306  1 24.8752 1.168e-05 ***
## a.fras      114.046  1 44.1163 5.290e-08 ***
## b.fras        9.239  1  3.5738    0.06578 .
## Residuals   105.990 41
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
pairs(cover[species=="Abies fraseri"] ~ a.fras + b.fras, data=trees)
```

```
#Again removing least sig term (b <- tci)...

abies_mod5 <- lm(cover[species=="Abies fraseri"] ~ a.fras, data=trees)
summary(abies_mod5) #Adj.R2=0.47
```

```
##
## Call:
## lm(formula = cover[species == "Abies fraseri"] ~ a.fras, data = trees)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -3.679 -1.488  0.488  1.214  3.418
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -15.814670   3.526227  -4.485 5.56e-05 ***
## a.fras        0.011914   0.001919   6.208 1.99e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.656 on 42 degrees of freedom
## Multiple R-squared:  0.4786, Adjusted R-squared:  0.4661
## F-statistic: 38.54 on 1 and 42 DF,  p-value: 1.991e-07
```

```
Anova(abies_mod5, type=3)
```

```
## Anova Table (Type III tests)
##
## Response: cover[species == "Abies fraseri"]
##              Sum Sq Df F value     Pr(>F)
## (Intercept)  55.184  1  20.114 5.561e-05 ***
## a.fras      105.749  1  38.545 1.991e-07 ***
## Residuals   115.229 42
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

#Although this final model (with only elev as an explanatory variable) has a slightly lower Adj.R2 value, it is also much less complex, and I think it could be argued that this is the best model. For Abies fraseri, the most important explnatory variable is elev. Overall, the models for the habitat specialist (Abies fraseri) do a much better job of explaining the variation in cover than the models for the habitat generalist (Acer rubrum). This is not surprising, given that the habitat specialist is a specialist for a reason, and it appears that elevation might be controlling for that. By definition, with a habitat generalist, there likely is not a single variable that can explain its distribution (or, in this case, cover).

#Checking to see if any assumptions are violated for Abies fraseri...
```
par(mfrow=c(2,2))
plot(abies_mod5)
```
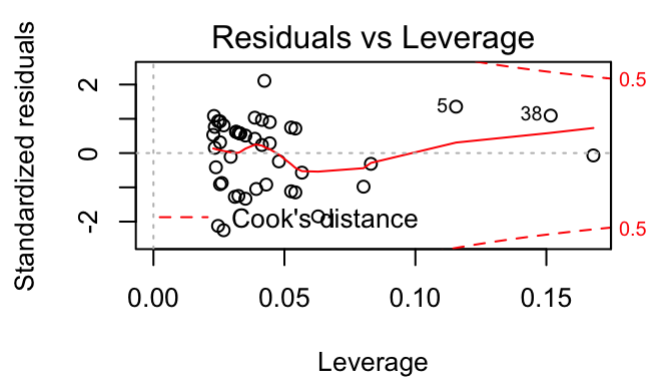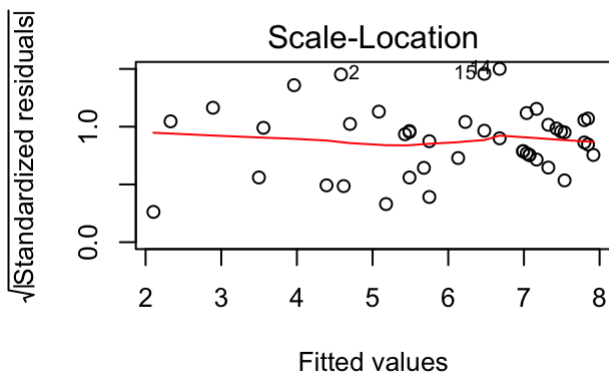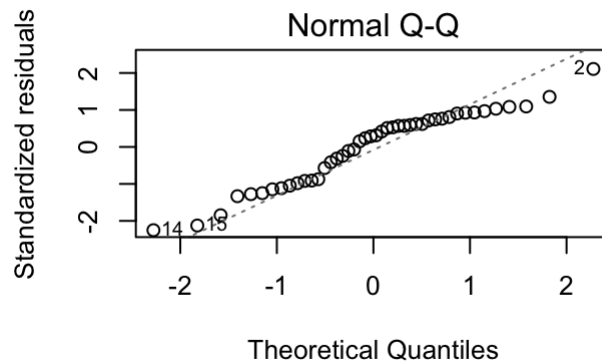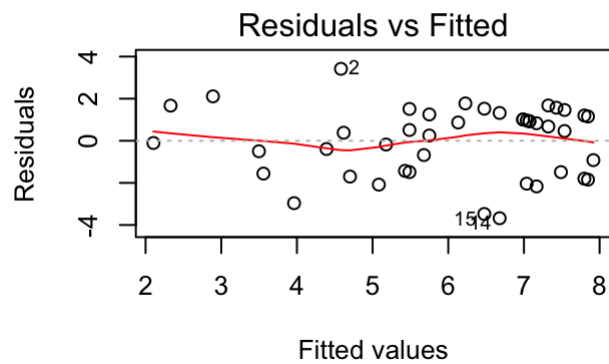
```
#2. Re-examining solutions to #1 with the perspective of a GLM...

#Best fitting model for Acer rubrum...
acer_modinteract <- lm(cover[species=="Acer rubrum"] ~ a + c + e + a*e, data=trees)
summary(acer_modinteract) #Adj.R2=0.061
```

```
##
## Call:
## lm(formula = cover[species == "Acer rubrum"] ~ a + c + e + a *
##     e, data = trees)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.9675 -1.2255  0.2472  1.3318  5.3015
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.3684250  0.3884493  11.246  < 2e-16 ***
## a            0.0008955  0.0004452   2.011 0.044651 *
## c            0.0017225  0.0004516   3.814 0.000149 ***
## e            1.0924924  0.3145520   3.473 0.000545 ***
## a:e         -0.0016507  0.0003505  -4.709 2.99e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.963 on 718 degrees of freedom
## Multiple R-squared:  0.06574,    Adjusted R-squared:  0.06053
## F-statistic: 12.63 on 4 and 718 DF,  p-value: 6.155e-10
```

```
acer_modinteract_glm <- glm(cover[species=="Acer rubrum"] ~ a + c + e + a*e, data=trees,
 family='poisson')
summary(acer_modinteract_glm)
```

```
##
## Call:
## glm(formula = cover[species == "Acer rubrum"] ~ a + c + e + a *
##      e, family = "poisson", data = trees)
##
## Deviance Residuals:
##     Min        1Q     Median        3Q       Max
## -2.5375  -0.5650    0.1061    0.5567    2.1915
##
## Coefficients:
##                 Estimate Std. Error z value Pr(>|z|)
## (Intercept)  1.484e+00  8.589e-02   17.280   < 2e-16 ***
## a            1.769e-04  9.702e-05    1.823 0.068245 .
## c            3.298e-04  9.932e-05    3.321 0.000898 ***
## e            2.207e-01  7.048e-02    3.132 0.001738 **
## a:e         -3.345e-04  7.936e-05   -4.215 2.49e-05 ***
## ---
## 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: 610.52  on 718  degrees of freedom
## AIC: 3082.9
##
## Number of Fisher Scoring iterations: 4
```

```
#Best fitting model for Abies fraseri...
abies_mod5 <- lm(cover[species=="Abies fraseri"] ~ a.fras, data=trees)
summary(abies_mod5) #Adj.R2=0.47
```

```
##
## Call:
## lm(formula = cover[species == "Abies fraseri"] ~ a.fras, data = trees)
##
## Residuals:
##     Min       1Q Median       3Q      Max
## -3.679  -1.488   0.488   1.214    3.418
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept) -15.814670   3.526227   -4.485 5.56e-05 ***
## a.fras        0.011914   0.001919    6.208 1.99e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.656 on 42 degrees of freedom
## Multiple R-squared:  0.4786, Adjusted R-squared:  0.4661
## F-statistic: 38.54 on 1 and 42 DF,  p-value: 1.991e-07
```

```
abies_mod5_glm <- glm(cover[species=="Abies fraseri"] ~ a.fras, data=trees, family='pois
son')
summary(abies_mod5_glm)
```

```
##
## Call:
## glm(formula = cover[species == "Abies fraseri"] ~ a.fras, family = "poisson",
##     data = trees)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.7627  -0.5757   0.2390   0.4400   1.5342
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept) -2.3784141  1.0143217  -2.345    0.019 *
## a.fras       0.0022556  0.0005425   4.158 3.21e-05 ***
## ---
## 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: 22.180  on 42  degrees of freedom
## AIC: 183.36
##
## Number of Fisher Scoring iterations: 4
```

```
pseudo_r2 = function(glm_mod) {
    1 -  glm_mod$deviance / glm_mod$null.deviance
  }

pseudo_r2(acer_modinteract_glm) #R2=0.060
```

```
## [1] 0.05978137
```

```
pseudo_r2(abies_mod5_glm) #R2=0.46
```

```
## [1] 0.4626207
```

```
#Comparing the residual sums of squares between the traditional models and the GLM model
s...
anova(acer_modinteract, acer_modinteract_glm)
```

```
## Analysis of Variance Table
##
## Model 1: cover[species == "Acer rubrum"] ~ a + c + e + a * e
## Model 2: cover[species == "Acer rubrum"] ~ a + c + e + a * e
##   Res.Df      RSS Df Sum of Sq F Pr(>F)
## 1     718 2766.59
## 2     718  610.52  0     2156.1
```

```
anova(abies_mod5, abies_mod5_glm)
```

```
## Analysis of Variance Table
##
## Model 1: cover[species == "Abies fraseri"] ~ a.fras
## Model 2: cover[species == "Abies fraseri"] ~ a.fras
##   Res.Df      RSS Df Sum of Sq F Pr(>F)
## 1     42 115.23
## 2     42  22.18  0     93.048
```

```
#Chaning the error distribution did not change the results much. Both models explain app
roximately the same amount of variation (R2=0.033 and R2=0.032). However, moving to the
 poission reduced the deviance dramatially for both species...
```

```
#3. Summary of results...

#For both species, I started with the most complex model (i.e. including all possible ex
planatory variables), and removed variables 1 by 1 (each time removing the variable with
 the least significance) until all variables were significant. Once I found a model with
 all significant variables, I began to rely more on the R2 value to find the best fittin
g model. In the case of Abies fraseri, the model with 2 explanatory variables (elev and
 tci) actually had a slightly higher AIC than the model with just 1 explanatory variable
 (elev). However, the difference in AIC was not much, and because a simpler model is alw
ays better, I chose the model with just 1 variable. I then plotted the final models to c
heck for any assumption violations. Overall, it is not surprising that the variation in
 cover for the habitat specialist can be better explained. It is also not surprising tha
t it could take more explanatory variables to explain the variation in cover for a habit
at generalist.
```

```
#4. Stepwise regression analysis on the full model...
library(MASS)

stepAIC(acer_mod1)
```

```
## Start:  AIC=1002.17
## cover[species == "Acer rubrum"] ~ a + b + c + as.factor(d) +
##     e
##
##                Df Sum of Sq    RSS     AIC
## - as.factor(d)  3     9.449 2837.7  998.58
## <none>                       2828.2 1002.17
## - b             1    12.581 2840.8 1003.37
## - c             1    29.085 2857.3 1007.56
## - e             1    35.613 2863.8 1009.21
## - a             1    40.439 2868.7 1010.43
##
## Step:  AIC=998.58
## cover[species == "Acer rubrum"] ~ a + b + c + e
##
##        Df Sum of Sq    RSS     AIC
## <none>             2837.7  998.58
## - b     1    14.370 2852.0 1000.23
## - c     1    31.491 2869.2 1004.56
## - e     1    35.515 2873.2 1005.57
## - a     1    45.778 2883.4 1008.15
```

```
##
## Call:
## lm(formula = cover[species == "Acer rubrum"] ~ a + b + c + e,
##     data = trees)
##
## Coefficients:
## (Intercept)            a            b            c            e
##   6.3218898   -0.0008868   -0.0668631    0.0013256   -0.3204370
```

```
#According to the stepwise AIC function, the best fitting model for Acer rubrum includes
 elev, tci, streamdist, and beers. This model has one more variable (tci) than my analys
is above.


stepAIC(abies_mod1)
```

```
## Start:  AIC=48.59
## cover[species == "Abies fraseri"] ~ a.fras + b.fras + c.fras +
##     as.factor(d.fras) + e.fras
##
##                     Df Sum of Sq     RSS     AIC
## - e.fras             1     0.014  92.304  46.599
## - as.factor(d.fras)  3    10.089 102.379  47.157
## - c.fras             1     1.636  93.926  47.366
## <none>                            92.289  48.593
## - b.fras             1     5.667  97.956  49.215
## - a.fras             1    61.618 153.908  69.095
##
## Step:  AIC=46.6
## cover[species == "Abies fraseri"] ~ a.fras + b.fras + c.fras +
##     as.factor(d.fras)
##
##                     Df Sum of Sq     RSS     AIC
## - c.fras             1     1.665  93.969 45.386
## - as.factor(d.fras)  3    10.679 102.983 45.417
## <none>                            92.304 46.599
## - b.fras             1     6.745  99.049 47.703
## - a.fras             1    64.662 156.966 67.961
##
## Step:  AIC=45.39
## cover[species == "Abies fraseri"] ~ a.fras + b.fras + as.factor(d.fras)
##
##                     Df Sum of Sq     RSS     AIC
## - as.factor(d.fras)  3    12.021 105.990 44.683
## <none>                            93.969 45.386
## - b.fras             1     6.807 100.776 46.463
## - a.fras             1    78.687 172.656 70.153
##
## Step:  AIC=44.68
## cover[species == "Abies fraseri"] ~ a.fras + b.fras
##
##           Df Sum of Sq    RSS     AIC
## <none>                  105.99 44.683
## - b.fras  1     9.239 115.23 46.360
## - a.fras  1   114.046 220.04 74.822
```

```
##
## Call:
## lm(formula = cover[species == "Abies fraseri"] ~ a.fras + b.fras,
##     data = trees)
##
## Coefficients:
## (Intercept)        a.fras        b.fras
##   -18.78984       0.01262       0.30454
```

```
#According to the stepwise AIC function, the best fitting model for Abies fraseri includ
es elev and tci. This model also has one more variable (tci) than my analysis above.
```