

## Assignment #3

Chelsea Woodruff  
2/4/2019

```
library(ggplot2)
library(gridExtra)
library(scatterplot3d)
library(MASS)
```

1.) To compare the variables stream distribution, elevation, topographic convergence index (tci), disturbance, and beers against cover in the Frasier Fir and Red Maple, respectively:

```
trees_abi <- subset(treedata, treedata$spcode == 'ABIEFRA')
trees_ace <- subset(treedata, treedata$spcode == 'ACERRUB')

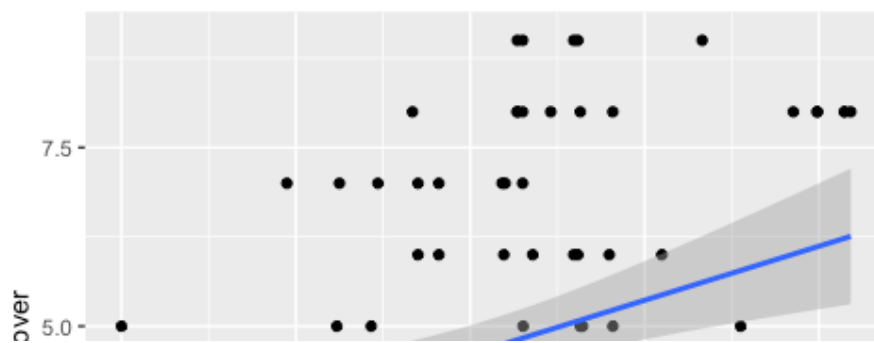
fir_big_lm <- lm(cover ~ streamdist + elev + tci + disturb + beers, data = trees_abi)
maple_big_lm <- lm(cover ~ streamdist + elev + tci + disturb + beers, data = trees_ace)
summary(fir_big_lm)
##
## Call:
## lm(formula = cover ~ streamdist + elev + tci + disturb + beers,
##     data = trees_abi)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.8668 -1.0262  0.1864  1.1427  4.2435
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -9.618273   2.110551  -4.557 1.56e-05 ***
## streamdist     0.001493   0.001070   1.395 0.16621
## elev           0.005447   0.001265   4.305 4.09e-05 ***
## tci            0.512971   0.142251   3.606 0.00050 ***
## disturbLT-SEL  0.336167   1.004582   0.335 0.73865
## disturbSETTLE  0.027070   1.298922   0.021 0.98342
## disturbVIRGIN  0.759546   0.580126   1.309 0.19363
## beers         0.780874   0.261330   2.988 0.00358 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.676 on 94 degrees of freedom
## Multiple R-squared:  0.4455, Adjusted R-squared:  0.4042
## F-statistic: 10.79 on 7 and 94 DF, p-value: 6.482e-10
```

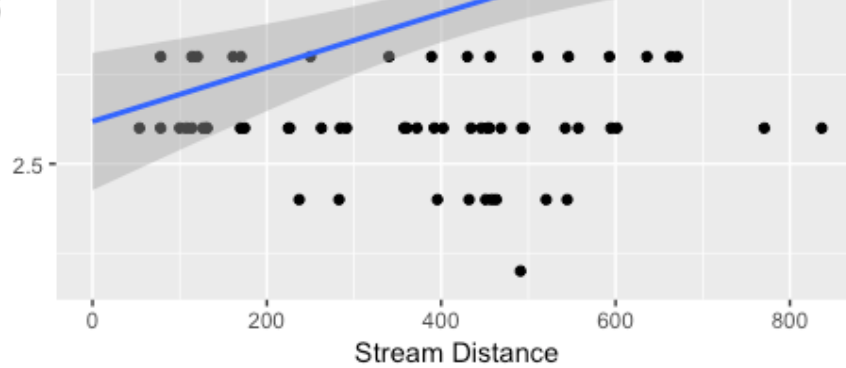
The table indicates that elevation, the topographic convergent index, and beers are all statistically significant in explaining coverage of the Frasier Fir tree. The p-values are all <0.05, indicating statistical significance.

Below, we will plot a few variables to see their individual effects on Frasier Fir coverage.

```
cover_streamdist <- ggplot(data = trees_abi,
  mapping = aes(x = streamdist, y = cover)) +
  geom_point() +
  geom_smooth(method = 'lm') +
  labs(x = 'Stream Distance', y = 'Cover')
```

cover\_streamdist

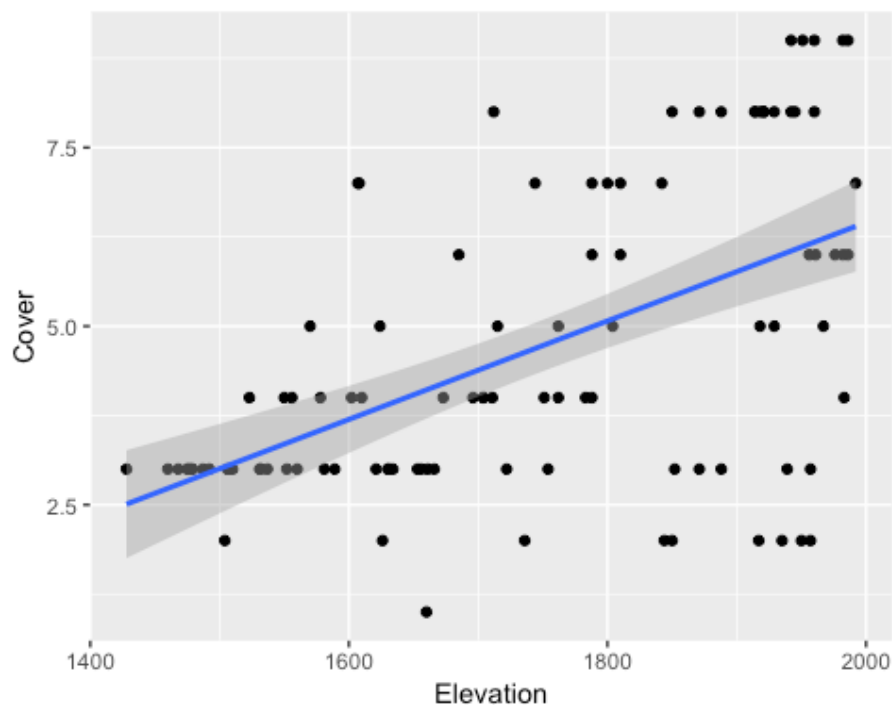




Above is the level coverage of the Frasier Fir tree in relation to distance from the closest stream water source. This is a linear model showing a line of best fit over the data points. Here we see the difficulty of plotting points with data sets containing a discrete variable (cover). While this variable of stream distance did not necessarily indicate statistical significance in the table above, below we can find the p-value to be 0.000456.

```
summary(lm(cover ~ streamdist, data = trees_abi))
##
## Call:
## lm(formula = cover ~ streamdist, data = trees_abi)
##
## Residuals:
##    Min     1Q   Median     3Q    Max
## -3.9495 -1.5940 -0.4854  1.7635  4.1900
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.092573   0.484243   6.386 5.39e-09 ***
## streamdist   0.003783   0.001043   3.625 0.000456 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.051 on 100 degrees of freedom
## Multiple R-squared:  0.1162, Adjusted R-squared:  0.1073
## F-statistic: 13.14 on 1 and 100 DF, p-value: 0.000456
abi_cover_elev <- ggplot(data = trees_abi,
  mapping = aes(x = elev, y = cover)) +
  geom_point() +
  geom_smooth(method = 'lm') +
  labs(x = 'Elevation', y = 'Cover')
```

abi\_cover\_elev

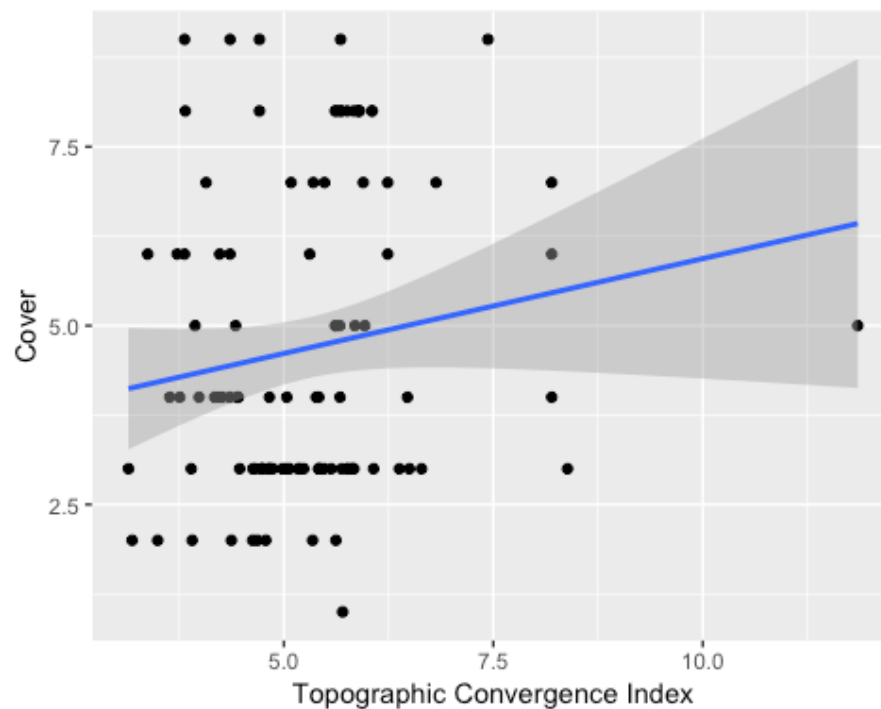


Here is a line of best fit placed over data explaining the relationship that variable elevation might have on the coverage of the Frasier Fir. By “squint test” we can see the line of best fit is slightly closer. But below, we have calculated the R-squared value to be 0.29. The p-

value of 3.122e-09 shows a statistically significant relationship between elevation and tree cover in the Frasier Fir, as was suggested by the first table.

```
summary(lm(cover ~ elev, data = trees_abi))
##
## Call:
## lm(formula = cover ~ elev, data = trees_abi)
##
## Residuals:
##   Min     1Q   Median     3Q      Max
## -4.1558 -0.9498 -0.0880  1.6055  3.5329
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -7.332833   1.856838  -3.949 0.000146 ***
## elev         0.006893   0.001060   6.504 3.12e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.829 on 100 degrees of freedom
## Multiple R-squared:  0.2973, Adjusted R-squared:  0.2902
## F-statistic: 42.3 on 1 and 100 DF, p-value: 3.122e-09
abi_cover_tci <- ggplot(data = trees_abi,
  mapping = aes(x = tci, y = cover)) +
  geom_point() +
  geom_smooth(method = 'lm') +
  labs(x = 'Topographic Convergence Index', y = 'Cover')
```

abi\_cover\_tci



The above plot is a good example of why examining large data sets by plotting is always helpful. Here we see a potential outlier on the far end of the plotted data. Might have been hard to catch without graphing. This also violates an OLS assumption of no error or outliers existing within the independent variable.

```
summary(lm(cover ~ tci, data = trees_abi))
##
## Call:
## lm(formula = cover ~ tci, data = trees_abi)
##
## Residuals:
##   Min     1Q   Median     3Q      Max
## -3.7971 -1.7092 -0.5929  1.7964  4.7031
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.2856    0.9403   3.494 0.00071 ***
## tci          0.2651    0.1733   1.529 0.12931
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```

##
## Residual standard error: 2.157 on 100 degrees of freedom
## Multiple R-squared: 0.02286, Adjusted R-squared: 0.01309
## F-statistic: 2.339 on 1 and 100 DF, p-value: 0.1293

```

Above is the tabular data of the individual TCI variable likely to explain coverage in the Fraser Fir. But here, the p-value does not indicate any statistical significance, as it did in the first table. This may be because of interaction between the variables. Below is a linear model assuming an all-interaction effect between the variables:

```

abi_int_mod <- lm(cover ~ elev * tci * streamdist * disturb * beers, data=trees_abi)
summary(abi_int_mod)

```

```

##
## Call:
## lm(formula = cover ~ elev * tci * streamdist * disturb * beers,
##     data = trees_abi)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.9037 -0.6839  0.0000  0.7603  4.2009
##
## Coefficients: (31 not defined because of singularities)
##              Estimate Std. Error t value
## (Intercept)   -1.647e+01  2.447e+02  -0.067
## elev          -2.728e-02  1.154e-01  -0.236
## tci           -5.639e+01  1.093e+02  -0.516
## streamdist     8.054e-01  1.361e+00   0.592
## disturbLT-SEL  3.058e+02  6.225e+02   0.491
## disturbSETTLE -1.461e+04  2.633e+04  -0.555
## disturbVIRGIN -3.663e+01  2.364e+02  -0.155
## beers         -5.727e+01  3.401e+02  -0.168
## elev:tci       4.380e-02  8.837e-02   0.496
## elev:streamdist -3.789e-04  6.092e-04  -0.622
## tci:streamdist -8.392e-02  8.908e-02  -0.942
## elev:disturbLT-SEL -1.246e-01  2.554e-01  -0.488
## elev:disturbSETTLE  8.343e+00  1.504e+01   0.555
## elev:disturbVIRGIN  6.416e-02  1.064e-01   0.603
## tci:disturbLT-SEL -1.806e+01  3.571e+01  -0.506
## tci:disturbSETTLE      NA         NA      NA
## tci:disturbVIRGIN  6.503e+01  1.086e+02   0.599
## streamdist:disturbLT-SEL -1.202e-02  5.907e-02  -0.203
## streamdist:disturbSETTLE      NA         NA      NA
## streamdist:disturbVIRGIN -5.813e-01  1.352e+00  -0.430
## elev:beers      -3.238e-02  1.905e-01  -0.170
## tci:beers       8.078e+00  2.068e+01   0.391
## streamdist:beers -2.097e-03  2.280e-01  -0.009
## disturbLT-SEL:beers      NA         NA      NA
## disturbSETTLE:beers      NA         NA      NA
## disturbVIRGIN:beers     7.679e+01  3.196e+02   0.240
## elev:tci:streamdist  2.491e-05  2.544e-05   0.979
## elev:tci:disturbLT-SEL      NA         NA      NA
## elev:tci:disturbSETTLE      NA         NA      NA
## elev:tci:disturbVIRGIN -4.951e-02  8.796e-02  -0.563
## elev:streamdist:disturbLT-SEL      NA         NA      NA
## elev:streamdist:disturbSETTLE      NA         NA      NA
## elev:streamdist:disturbVIRGIN  2.358e-04  6.000e-04   0.393
## tci:streamdist:disturbLT-SEL      NA         NA      NA
## tci:streamdist:disturbSETTLE      NA         NA      NA
## tci:streamdist:disturbVIRGIN  4.529e-02  8.020e-02   0.565
## elev:tci:beers     2.986e-03  9.268e-03   0.322
## elev:streamdist:beers  6.966e-05  1.105e-04   0.630
## tci:streamdist:beers  1.577e-02  3.514e-02   0.449
## elev:disturbLT-SEL:beers      NA         NA      NA
## elev:disturbSETTLE:beers      NA         NA      NA
## elev:disturbVIRGIN:beers  1.699e-02  1.774e-01   0.096
## tci:disturbLT-SEL:beers      NA         NA      NA
## tci:disturbSETTLE:beers      NA         NA      NA
## tci:disturbVIRGIN:beers -1.178e+01  1.377e+01  -0.856
## streamdist:disturbLT-SEL:beers      NA         NA      NA
## streamdist:disturbSETTLE:beers      NA         NA      NA
## streamdist:disturbVIRGIN:beers -9.767e-02  1.319e-01  -0.741
## elev:tci:streamdist:disturbLT-SEL      NA         NA      NA
## elev:tci:streamdist:disturbSETTLE      NA         NA      NA
## elev:tci:streamdist:disturbVIRGIN      NA         NA      NA
## elev:tci:streamdist:beers -1.143e-05  2.123e-05  -0.538
## elev:tci:streamdist:disturbLT-SEL:beers      NA         NA      NA

```

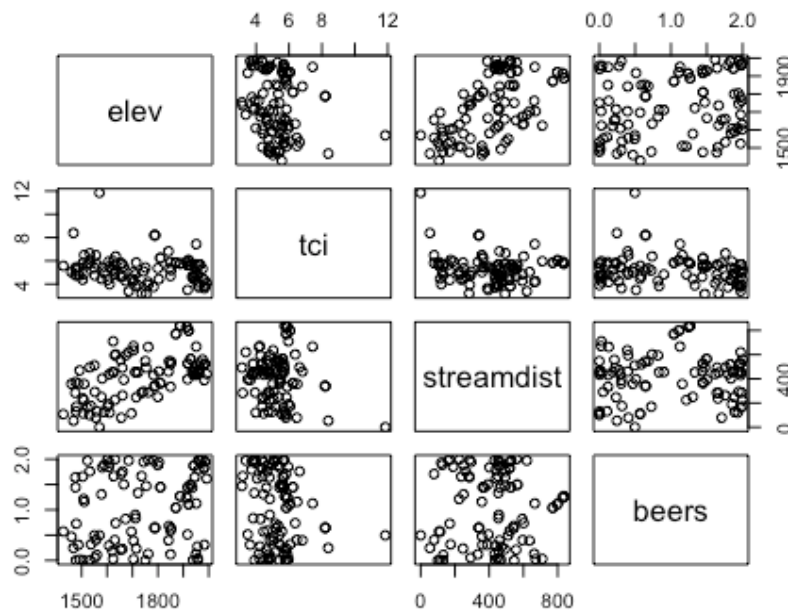
## elev:tc1:disturbLT-SEL:beers	NA	NA	NA
## elev:tc1:disturbVIRGIN:beers	NA	NA	NA
## elev:streamdist:disturbLT-SEL:beers	NA	NA	NA
## elev:streamdist:disturbSETTLE:beers	NA	NA	NA
## elev:streamdist:disturbVIRGIN:beers	NA	NA	NA
## tc1:streamdist:disturbLT-SEL:beers	NA	NA	NA
## tc1:streamdist:disturbSETTLE:beers	NA	NA	NA
## tc1:streamdist:disturbVIRGIN:beers	NA	NA	NA
## elev:tc1:streamdist:disturbLT-SEL:beers	NA	NA	NA
## elev:tc1:streamdist:disturbSETTLE:beers	NA	NA	NA
## elev:tc1:streamdist:disturbVIRGIN:beers	NA	NA	NA
##	Pr(>lt)		
## (Intercept)	0.947		
## elev	0.814		
## tc1	0.607		
## streamdist	0.556		
## disturbLT-SEL	0.625		
## disturbSETTLE	0.581		
## disturbVIRGIN	0.877		
## beers	0.867		
## elev:tc1	0.622		
## elev:streamdist	0.536		
## tc1:streamdist	0.349		
## elev:disturbLT-SEL	0.627		
## elev:disturbSETTLE	0.581		
## elev:disturbVIRGIN	0.548		
## tc1:disturbLT-SEL	0.615		
## tc1:disturbSETTLE	NA		
## tc1:disturbVIRGIN	0.551		
## streamdist:disturbLT-SEL	0.839		
## streamdist:disturbSETTLE	NA		
## streamdist:disturbVIRGIN	0.669		
## elev:beers	0.866		
## tc1:beers	0.697		
## streamdist:beers	0.993		
## disturbLT-SEL:beers	NA		
## disturbSETTLE:beers	NA		
## disturbVIRGIN:beers	0.811		
## elev:tc1:streamdist	0.331		
## elev:tc1:disturbLT-SEL	NA		
## elev:tc1:disturbSETTLE	NA		
## elev:tc1:disturbVIRGIN	0.575		
## elev:streamdist:disturbLT-SEL	NA		
## elev:streamdist:disturbSETTLE	NA		
## elev:streamdist:disturbVIRGIN	0.696		
## tc1:streamdist:disturbLT-SEL	NA		
## tc1:streamdist:disturbSETTLE	NA		
## tc1:streamdist:disturbVIRGIN	0.574		
## elev:tc1:beers	0.748		
## elev:streamdist:beers	0.531		
## tc1:streamdist:beers	0.655		
## elev:disturbLT-SEL:beers	NA		
## elev:disturbSETTLE:beers	NA		
## elev:disturbVIRGIN:beers	0.924		
## tc1:disturbLT-SEL:beers	NA		
## tc1:disturbSETTLE:beers	NA		
## tc1:disturbVIRGIN:beers	0.395		
## streamdist:disturbLT-SEL:beers	NA		
## streamdist:disturbSETTLE:beers	NA		
## streamdist:disturbVIRGIN:beers	0.461		
## elev:tc1:streamdist:disturbLT-SEL	NA		
## elev:tc1:streamdist:disturbSETTLE	NA		
## elev:tc1:streamdist:disturbVIRGIN	NA		
## elev:tc1:streamdist:beers	0.592		
## elev:tc1:disturbLT-SEL:beers	NA		
## elev:tc1:disturbSETTLE:beers	NA		
## elev:tc1:disturbVIRGIN:beers	NA		
## elev:streamdist:disturbLT-SEL:beers	NA		
## elev:streamdist:disturbSETTLE:beers	NA		
## elev:streamdist:disturbVIRGIN:beers	NA		
## tc1:streamdist:disturbLT-SEL:beers	NA		
## tc1:streamdist:disturbSETTLE:beers	NA		
## tc1:streamdist:disturbVIRGIN:beers	NA		
## elev:tc1:streamdist:disturbLT-SEL:beers	NA		
## elev:tc1:streamdist:disturbSETTLE:beers	NA		

```
## elev:tci:streamdist:disturbVIRGIN:beers    NA
##
## Residual standard error: 1.684 on 69 degrees of freedom
## Multiple R-squared:  0.5888, Adjusted R-squared:  0.3981
## F-statistic: 3.088 on 32 and 69 DF, p-value: 4.515e-05
```

To explain the variation of the p-values from being run individually or jointly, we must assume these explanatory variables are not entirely independent from one another. Above is the result if we assume an all interaction model. By first glance, we can assume this is not a useful model.

To test the variables individual interaction upon one another, we use the 'pairs' function below:

```
pairs(~ elev + tci + streamdist + beers, data = trees_abi)
```



```
lm(cover ~ elev + tci + streamdist + beers + elev:tci + elev:streamdist + elev:beers + tci:streamdist
+ tci:beers + streamdist:beers + elev:tci:streamdist:beers, data = trees_abi)
##
## Call:
## lm(formula = cover ~ elev + tci + streamdist + beers + elev:tci +
##   elev:streamdist + elev:beers + tci:streamdist + tci:beers +
##   streamdist:beers + elev:tci:streamdist:beers, data = trees_abi)
##
## Coefficients:
##      (Intercept)          elev
##      2.972e+01      -1.695e-02
##           tci      streamdist
##      -4.224e+00      -1.954e-02
##           beers      elev:tci
##      -1.175e+01      2.703e-03
##      elev:streamdist      elev:beers
##       1.128e-05      6.347e-03
##       tci:streamdist      tci:beers
##       4.229e-04      4.139e-01
##      streamdist:beers      elev:tci:streamdist:beers
##       5.331e-03      -7.743e-07
```

By using the information we have found from all the models and tables above, we can begin to create a more accurate model for our data.

```
new_abi_lm <- lm(cover ~ tci + elev + beers, data = trees_abi)
summary(new_abi_lm)
##
## Call:
## lm(formula = cover ~ tci + elev + beers, data = trees_abi)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.9474 -1.2389  0.0673  1.1258  4.4466
```

```
##
## Coefficients:
##      Estimate Std. Error t value Pr(>|t|)
## (Intercept) -10.426571  1.965465  -5.305 7.00e-07 ***
## tci          0.505556  0.137752   3.670 0.000395 ***
## elev         0.006623  0.001005   6.587 2.25e-09 ***
## beers        0.848679  0.245905   3.451 0.000825 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.67 on 98 degrees of freedom
## Multiple R-squared:  0.4257, Adjusted R-squared:  0.4081
## F-statistic: 24.21 on 3 and 98 DF, p-value: 8.319e-12
```

The above model indicates that the rate of tree coverage in the Fraser Fir is dependent on the TCI, elevation, and beers. These variables are the most important when predicting the natural rate of coverage.

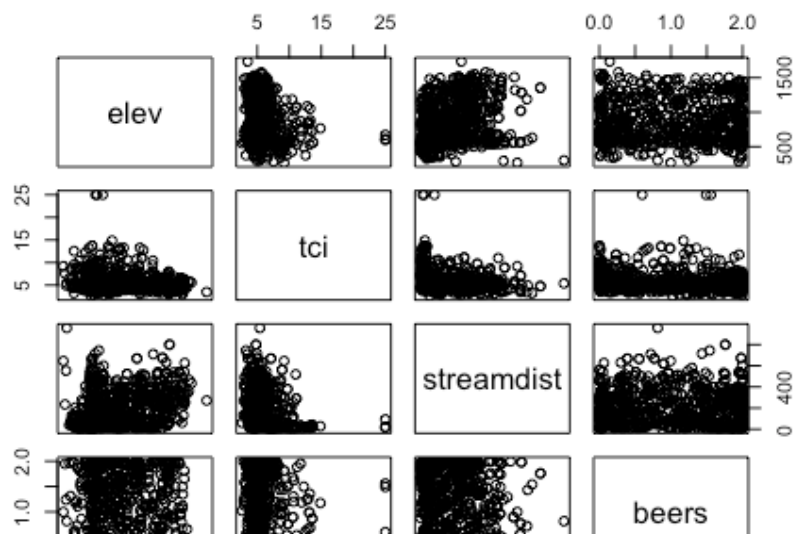
Below we can view the summary of our next data set, the Red Maple.

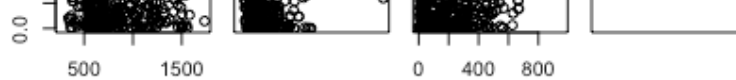
```
summary(maple_big_lm)
##
## Call:
## lm(formula = cover ~ streamdist + elev + tci + disturb + beers,
##     data = trees_ace)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -4.8054 -1.2410  0.3613  1.4124  5.1442
##
## Coefficients:
##      Estimate Std. Error t value Pr(>|t|)
## (Intercept)  6.0776176  0.4347601  13.979 < 2e-16 ***
## streamdist    0.0011543  0.0004597   2.511  0.01224 *
## elev        -0.0007668  0.0002960  -2.590  0.00977 **
## tci         -0.0590939  0.0330850  -1.786  0.07446 .
## disturbLT-SEL 0.1865060  0.2096571   0.890  0.37397
## disturbSETTLE 0.1963689  0.2605137   0.754  0.45121
## disturbVIRGIN 0.3431481  0.2419342   1.418  0.15648
## beers       -0.3193888  0.1044973  -3.056  0.00232 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.013 on 786 degrees of freedom
## (1 observation deleted due to missingness)
## Multiple R-squared:  0.03802, Adjusted R-squared:  0.02945
## F-statistic: 4.438 on 7 and 786 DF, p-value: 7.48e-05
```

The model indicates that stream distribution, elevation, and beers are all statistically significant in explaining coverage of the Red Maple.

But, this is assuming no interaction is happening between the variables. Again, we will use the 'pairs' function to analyze the variables for interaction between each other.

```
pairs(~ elev+tcI+streamdist+beers, data=trees_ace)
```





```
lm(cover ~ elev + tci + streamdist + beers + elev:tci + elev:streamdist + elev:beers + tci:streamdist
+ tci:beers + streamdist:beers + elev:tci:streamdist:beers, data = trees_ace)
##
## Call:
## lm(formula = cover ~ elev + tci + streamdist + beers + elev:tci +
##   elev:streamdist + elev:beers + tci:streamdist + tci:beers +
##   streamdist:beers + elev:tci:streamdist:beers, data = trees_ace)
##
## Coefficients:
##      (Intercept)          elev
##      4.973e+00      8.287e-04
##          tci      streamdist
##      1.321e-02     -3.132e-03
##          beers      elev:tci
##      3.608e-01     -9.144e-05
##   elev:streamdist     elev:beers
##      2.099e-06     -9.399e-04
##   tci:streamdist      tci:beers
##      3.815e-04      8.686e-03
##   streamdist:beers elev:tci:streamdist:beers
##      3.309e-03     -5.841e-07
new_ace_lm <- lm(cover ~ tci + elev + streamdist + beers, data = trees_ace)
summary(new_ace_lm)
##
## Call:
## lm(formula = cover ~ tci + elev + streamdist + beers, data = trees_ace)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -4.8018 -1.2919  0.3624  1.4084  5.1641
##
## Coefficients:
##      Estimate Std. Error t value Pr(>|t|)
## (Intercept)  6.2836538  0.3441316  18.259 < 2e-16 ***
## tci         -0.0613091  0.0329629  -1.860  0.06327 .
## elev         -0.0007886  0.0002448  -3.222  0.00132 **
## streamdist   0.0011970  0.0004528   2.643  0.00837 **
## beers        -0.3267810  0.1028508  -3.177  0.00154 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.011 on 789 degrees of freedom
## (1 observation deleted due to missingness)
## Multiple R-squared:  0.03547, Adjusted R-squared:  0.03058
## F-statistic: 7.254 on 4 and 789 DF, p-value: 9.74e-06
```

The above model indicates that the rate of tree coverage in the Red Maple is dependent on the TCI, elevation, stream distance, and beers. These variables are the most important when predicting the natural rate of coverage.

OLS assumptions were checked between models and tables across both species. Linearity between coverage and independent variables was prevalent. Few errors or outliers were seen within independent variables with some exception, for example, to the elevation variable on coverage for the Fraser Fir species. Error around the regression line, or homoscedasticity, was difficult to measure by “squint test” but likely non-problematic in this case. Direction and magnitude of error between variables is likely seen between variables that exhibit interaction, as these do.

Differences in variance between the species can most likely be explained via the difference in the dataset size between species, as the Red Maple has more collected data points.

2.) Compare your qualitative assessment of which variables were most important in each model. Does it appear that changing the error distribution changed the results much? In what ways?

```
acer_poi = glm(cover ~ tci + elev + streamdist + disturb + beers, data = trees_ace,
family='poisson')
abi_poi = glm(cover ~ tci + elev + streamdist + disturb + beers, data = trees_abi, family='poisson')
pseudo_r2 = function(glm_mod) {1 - glm_mod$deviance / glm_mod$null.deviance}
pseudo_r2(abi_poi)
```



```
## [1] 0.4603601
#0.4603601
pseudo_r2(acer_poi)
## [1] 0.03385888
#0.03385888
```

```
summary(fir_big_lm) Multiple R-squared: 0.4455, Adjusted R-squared: 0.4042 p-value: 6.482e-10
```

```
summary(maple_big_lm) Multiple R-squared: 0.03802, Adjusted R-squared: 0.02945 p-value: 7.48e-05
```

Here we are providing the error term instead of asking the function to compute for us. The R-squared values were near identical with the change in error distribution. By predicting and then telling the R-squared value, the function ran almost the same as it had when we had asked it to compute that term for us, with minor differences.

3.) Provide a plain English summary (i.e., no statistics) of what you have found and what conclusions we can take away from your analysis?

When first analyzing a large dataset such as this one, it is important to get a general understanding of any relationship between variables. By experimenting with creating model and isolating independent variables in relation to the dependent, we can begin to see relationships. Running an ANOVA was key to pinpointing any statistical significance within the variables. But this was done assuming no interaction was taking effect by the variables. Because of this, the next step was to run a pairs analysis in order to visually view any interaction between variables. This allowed us to narrow down our focus of variables to just the statistically important few. The created models from there indicate that stream distribution, elevation, and beers are all statistically significant in explaining coverage of the Red Maple. And the additional model indicates that the rate of tree coverage in the Red Maple is dependent on the TCI, elevation, stream distance, and beers. These variables are the most important when predicting the natural rate of coverage. This might be important when trying to predict changes to the species' when changing the independent variables. For example, climate change might alter stream distances in the Smokey Mountains and scientists will want to predict the effect this might have on the Red Maple.

4.) (optional) Examine the behavior of the function stepAIC() using the exploratory models developed above. This is a very simple and not very robust machine learning stepwise algorithm that uses AIC to select a best model. By default it does a backward selection routine.

```
abi_step <- stepAIC(abi_poi, trace = FALSE)
```

```
View(abi_step)
```

```
abi_step$anova
```

Stepwise Model Path Analysis of Deviance Table

Initial Model: cover ~ tci + elev + streamdist + disturb + beers

Final Model: cover ~ tci + elev + beers

	Step	Df	Deviance	Resid. Df	Resid. Dev	AIC
1	94	53.32828	407.8549	2	- disturb	3 0.9748869 97 54.30317 402.8298
	3	- streamdist	1 0.6848549 98 54.98802 401.5146			

Here the function has eliminated variables it deemed unfit for the analysis.

```
acer_step <- stepAIC(acer_poi, trace = FALSE)
```

```
acer_step$anova
```

Stepwise Model Path Analysis of Deviance Table

Initial Model: cover ~ tci + elev + streamdist + disturb + beers

Final Model: cover ~ tci + elev + streamdist + beers

	Step	Df	Deviance	Resid. Df	Resid. Dev	AIC
--	------	----	----------	-----------	------------	-----

1 786 697.0335 3422.127 2 - disturb 3 1.742863 789 698.7764 3417.870

Both species' models in this case decided to include TCI, elevation, and beers. The other variables were deemed insignificant for the model.

5.) (optional) Develop a model for the number of species in each site (i.e., unique plotID). This variable will also be discrete so the Poisson may be a good starting approximation. Side note: the Poisson distribution converges asymptotically on the Gaussian distribution as the mean of the distribution increases. Thus Poisson regression does not differ much from traditional OLS when means are large.