Lab5

Joseph Crockett ES 207: Environmental Data Analysis March 1, 2016 Homework Assignment 4

Joseph Crockett ES 207: Environmental Data Analysis  
March 14th 2016  
Homework Assignment 5

Objective Statement:  
Preliminary results indicate that a linear function relates tree height to diameter at breast height (DBH) and genus. We will extend this analysis to include prediction of tree height based on DBH measurements, genus, and an additional environmental covariate to assess standing carbon at a project site without tree height measurements.

Methods:  
All analysis was conducted using R version 3.2.3 "Wooden Christmas Tree" (R Core Team 2015). We extracted mean August temperature, August precipitation, elevation, longitude, and latitude of our four project sites using the R packages rgdal and raster (Bivand 2015 and Hijmans 2015). After exploratory data analysis, we developed four models that took the form

In which ht is tree height, dbh is Diameter at Breast Height, genus is the five genera of greatest frequency at project sites, and parameter is temperature, precipitation, elevation, or latitude. We forced a zero intercept for each model to accurately depict the 0 - 0 relationship between height and dbh at germination. Each model was compared against the base interaction model ht ~ dbh \* genus and each other. P-values, R2, and AIC values were considered. The R function stepAIC from the MASS package was used to compare AIC values. Data:  
The data are tree measurements from 4 riparian project areas in Northern California. We have previously cleansed the dataset for illogical or unknown values and subsetted for the 5 most frequent genera. Outliers have been identified and removed. We extracted longitude, latitude, elevation, August precipitation, and mean Augus temperature from the raster files "DEM.tif", "precip\_8.tif", and "tmean\_8.tif".

Code:

# Load cleaned tree data and use raster and readGDAL to load DEM, precip\_8, and tmean\_8.  
load("~/Desktop/General/Spring 2016/ES 207/Labs/Projects/Lab-1/Data/rd\_lab3.rda")  
ProjLoc <- aggregate(cbind(Longitude,Latitude) ~ ProjCode,  
data=ripdata\_ss, mean)  
  
gdal\_grid2 = readGDAL("~/Desktop/General/Spring 2016/ES 207/Labs/Projects/Lab-1/Data/precip\_8.tif")

## ~/Desktop/General/Spring 2016/ES 207/Labs/Projects/Lab-1/Data/precip\_8.tif has GDAL driver GTiff   
## and has 862 rows and 744 columns

r2 = raster(gdal\_grid2)  
  
gdal\_grid1 = readGDAL("~/Desktop/General/Spring 2016/ES 207/Labs/Projects/Lab-1/Data/DEM.tif")

## ~/Desktop/General/Spring 2016/ES 207/Labs/Projects/Lab-1/Data/DEM.tif has GDAL driver GTiff   
## and has 1137 rows and 1233 columns

r1 = raster(gdal\_grid1)  
  
gdal\_grid3 = readGDAL("~/Desktop/General/Spring 2016/ES 207/Labs/Projects/Lab-1/Data/tmean\_8.tif")

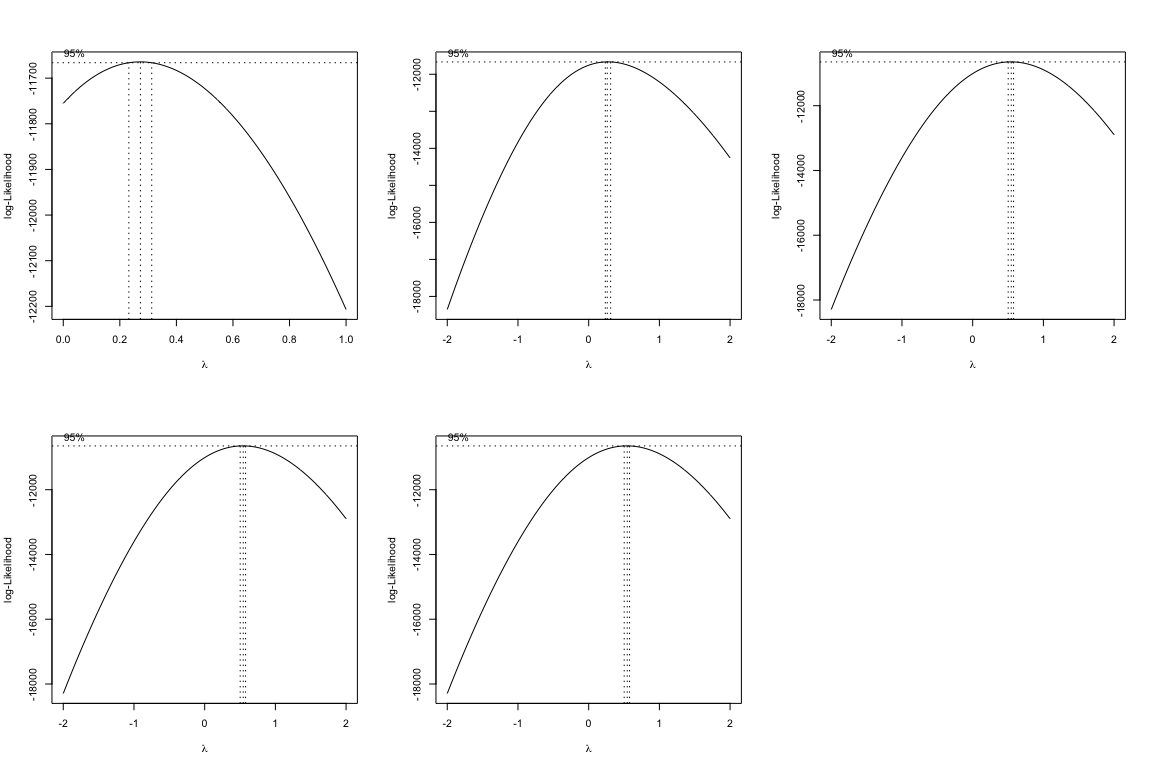
## ~/Desktop/General/Spring 2016/ES 207/Labs/Projects/Lab-1/Data/tmean\_8.tif has GDAL driver GTiff   
## and has 1137 rows and 1233 columns

r3 = raster(gdal\_grid3)

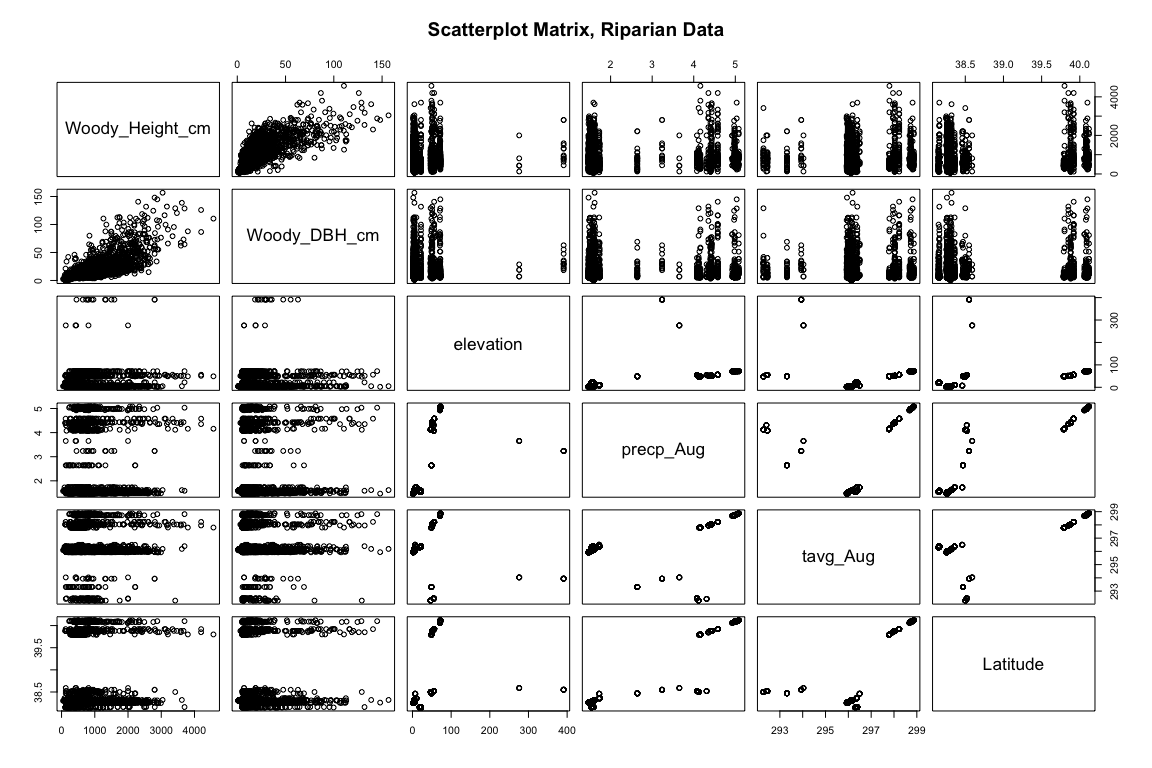
# Plot the precip data using ggplot 2 and superimpose project locations on the map.  
r2\_p <- data.frame(rasterToPoints(r2))  
colnames(r2\_p) <- c("Longitude", "Latitude", "Precip")  
  
r2\_pp <- ggplot(data = r2\_p, aes(x = Longitude, y = Latitude)) + geom\_raster(aes(fill = Precip)) + geom\_point(data = ProjLoc, aes(x = Longitude, y = Latitude, color = ProjCode), size = 3, shape = 4) + theme\_minimal() + coord\_equal() + scale\_fill\_gradient("August Precipitation(mm)", limits = c(-1, 25))  
  
#DEM  
r1\_p <- data.frame(rasterToPoints(r1))  
colnames(r1\_p) <- c("Longitude", "Latitude", "DEM")  
r1\_pf <- filter(r1\_p, Longitude >= min(r2\_p$Longitude) & Longitude <= max(r2\_p$Longitude) & Latitude >= min(r2\_p$Latitude) & Latitude <= max(r2\_p$Latitude))  
  
r1\_pp<- ggplot(data = r1\_pf, aes(x = Longitude, y = Latitude)) + geom\_raster(aes(fill = DEM)) + geom\_point(data = ProjLoc, aes(x = Longitude, y = Latitude, color = ProjCode), size = 3, shape = 4) + theme\_minimal() + coord\_equal() + scale\_fill\_gradient("DEM (ft)", low = "black", high = "white",limits = c(-90, 4000)) + theme(legend.justification=c(1,0), legend.position=c(0,1))  
  
#tavg  
r3\_p <- data.frame(rasterToPoints(r3))  
colnames(r3\_p) <- c("Longitude", "Latitude", "Tavg")  
r3\_pf <- filter(r3\_p, Longitude >= min(r2\_p$Longitude) & Longitude <= max(r2\_p$Longitude) & Latitude >= min(r2\_p$Latitude) & Latitude <= max(r2\_p$Latitude))  
  
r3\_pp <- ggplot(data = r3\_pf, aes(x = Longitude, y = Latitude)) + geom\_raster(aes(fill = Tavg)) + geom\_point(data = ProjLoc, aes(x = Longitude, y = Latitude, color = ProjCode), size = 3, shape = 4) + theme\_minimal() + coord\_equal() + scale\_fill\_gradient("August Temperature(K)", low = "gold", high = "darkred", limits = c(280, 310))  
  
# Extract values from rasters and add to data frame  
  
lat <- ripdata\_ss$Latitude  
lon <- ripdata\_ss$Longitude  
xy <- cbind(lon, lat)  
  
ripdata\_ss$precp\_Aug <- extract(r2, xy)  
ripdata\_ss$elevation <- extract(r1, xy)  
ripdata\_ss$tavg\_Aug <- extract(r3, xy)  
  
describe(ripdata\_ss) #pysch package

## vars n mean sd median trimmed mad  
## SurveyID 1 3230 346.02 365.92 82.00 324.70 83.03  
## ProjectID\* 2 3230 NaN NA NA NaN NA  
## LocationName\* 3 3230 NaN NA NA NaN NA  
## Date\* 4 3230 NaN NA NA NaN NA  
## Collectors\* 5 3230 NaN NA NA NaN NA  
## Longitude 6 3230 -121.54 0.35 -121.40 -121.48 0.03  
## Latitude 7 3230 38.55 0.62 38.27 38.41 0.01  
## SurveyTypeID\* 8 3230 NaN NA NA NaN NA  
## Plot.Name\* 9 3230 NaN NA NA NaN NA  
## SpeciesVarietalCode\* 10 3230 NaN NA NA NaN NA  
## SpeciesVarietalName\* 11 3230 NaN NA NA NaN NA  
## Measurement 12 3230 25.37 23.91 17.00 21.41 16.31  
## CanopyID\* 13 3230 NaN NA NA NaN NA  
## Woody\_DBH\_cm 14 3230 17.77 18.46 11.20 13.85 7.56  
## Woody\_Height\_m 15 3230 9.28 5.86 8.00 8.53 4.31  
## ProjCode\* 16 3230 NaN NA NA NaN NA  
## Genus\* 17 3230 NaN NA NA NaN NA  
## Woody\_Height\_cm 18 3230 927.93 585.58 800.00 852.53 431.44  
## precp\_Aug 19 3230 2.11 1.19 1.52 1.85 0.04  
## elevation 20 3230 17.94 33.55 5.00 11.77 2.97  
## tavg\_Aug 21 3230 296.32 1.08 296.01 296.22 0.07  
## min max range skew kurtosis se  
## SurveyID 6.00 857.00 851.00 0.44 -1.77 6.44  
## ProjectID\* Inf -Inf -Inf NA NA NA  
## LocationName\* Inf -Inf -Inf NA NA NA  
## Date\* Inf -Inf -Inf NA NA NA  
## Collectors\* Inf -Inf -Inf NA NA NA  
## Longitude -122.90 -121.19 1.71 -1.92 3.26 0.01  
## Latitude 38.15 40.12 1.97 1.82 1.40 0.01  
## SurveyTypeID\* Inf -Inf -Inf NA NA NA  
## Plot.Name\* Inf -Inf -Inf NA NA NA  
## SpeciesVarietalCode\* Inf -Inf -Inf NA NA NA  
## SpeciesVarietalName\* Inf -Inf -Inf NA NA NA  
## Measurement 1.00 152.00 151.00 1.44 1.73 0.42  
## CanopyID\* Inf -Inf -Inf NA NA NA  
## Woody\_DBH\_cm 0.90 156.70 155.80 3.01 11.56 0.32  
## Woody\_Height\_m 0.45 45.70 45.25 1.45 2.89 0.10  
## ProjCode\* Inf -Inf -Inf NA NA NA  
## Genus\* Inf -Inf -Inf NA NA NA  
## Woody\_Height\_cm 45.00 4570.00 4525.00 1.45 2.89 10.30  
## precp\_Aug 1.46 5.08 3.62 1.65 0.87 0.02  
## elevation 2.00 391.00 389.00 6.57 64.49 0.59  
## tavg\_Aug 292.28 298.87 6.59 -0.11 3.67 0.02

#Note: elevation, dbh, height, and precp are skewed right, while tavg is slightly skewed left. SD and Mean seem to scale together. A log transform worked well in the past, but perhaps a boxcox transformation would work as well.  
par(mfrow = c(2,3))  
boxcox(lm(Woody\_Height\_cm ~ precp\_Aug, data = ripdata\_ss), seq(0,1,0.1))  
boxcox(lm(Woody\_Height\_cm ~ Latitude, data = ripdata\_ss))  
boxcox(lm(Woody\_Height\_cm ~ Woody\_DBH\_cm, data = ripdata\_ss))  
boxcox(lm(Woody\_Height\_cm ~ Woody\_DBH\_cm, data = ripdata\_ss))  
boxcox(lm(Woody\_Height\_cm ~ Woody\_DBH\_cm, data = ripdata\_ss))  
par(mfrow = c(1,1))



prg <- stack(select(ripdata\_ss, elevation, precp\_Aug, tavg\_Aug, Latitude, Woody\_DBH\_cm, Woody\_Height\_cm))  
  
nw <- ggplot(prg, aes(x = values)) + geom\_density(aes(fill = ind), alpha = .3) + facet\_wrap(~ind, scales = "free", nrow = 5) + theme\_minimal() + guides(fill = F)  
# EDA on new values, specifically distribution and correlation  
  
pairs(~Woody\_Height\_cm + Woody\_DBH\_cm + elevation + precp\_Aug + tavg\_Aug + Latitude, data = ripdata\_ss, main = "Scatterplot Matrix, Riparian Data")



cor.test(~Woody\_Height\_cm + Woody\_DBH\_cm, data = ripdata\_ss)

##   
## Pearson's product-moment correlation  
##   
## data: Woody\_Height\_cm and Woody\_DBH\_cm  
## t = 63.787, df = 3228, p-value < 2.2e-16  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## 0.7310717 0.7616068  
## sample estimates:  
## cor   
## 0.7467324

cor.test(~Woody\_Height\_cm + elevation, data = ripdata\_ss)

##   
## Pearson's product-moment correlation  
##   
## data: Woody\_Height\_cm and elevation  
## t = 0.32087, df = 3228, p-value = 0.7483  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## -0.02884677 0.04012835  
## sample estimates:  
## cor   
## 0.00564751

cor.test(~Woody\_Height\_cm + precp\_Aug, data = ripdata\_ss)

##   
## Pearson's product-moment correlation  
##   
## data: Woody\_Height\_cm and precp\_Aug  
## t = -0.054854, df = 3228, p-value = 0.9563  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## -0.03545295 0.03352430  
## sample estimates:  
## cor   
## -0.000965472

cor.test(~Woody\_Height\_cm + tavg\_Aug, data = ripdata\_ss)

##   
## Pearson's product-moment correlation  
##   
## data: Woody\_Height\_cm and tavg\_Aug  
## t = 0.43556, df = 3228, p-value = 0.6632  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## -0.02682982 0.04214345  
## sample estimates:  
## cor   
## 0.007665932

cor.test(~Woody\_Height\_cm + Latitude, data = ripdata\_ss)

##   
## Pearson's product-moment correlation  
##   
## data: Woody\_Height\_cm and Latitude  
## t = 0.51414, df = 3228, p-value = 0.6072  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## -0.02544774 0.04352394  
## sample estimates:  
## cor   
## 0.009048863

ripdata\_ss$log\_DBH <- log(ripdata\_ss$Woody\_DBH\_cm)  
ripdata\_ss$log\_ht <- log(ripdata\_ss$Woody\_Height\_cm)  
ripdata\_ss$log\_elev <- sqrt(ripdata\_ss$elevation)  
ripdata\_ss$log\_pA <- log(ripdata\_ss$precp\_Aug)  
  
prg2 <- stack(select(ripdata\_ss, log\_DBH, log\_ht, tavg\_Aug, log\_elev, log\_pA, Latitude))  
  
tru <- ggplot(prg2, aes(x = values)) + geom\_density(aes(fill = ind), alpha = .3) + facet\_wrap(~ind, scales = "free", nrow = 5) + theme\_minimal() + guides(fill = F)  
  
#The log transformations do not appear to affect elevation or precipitation.

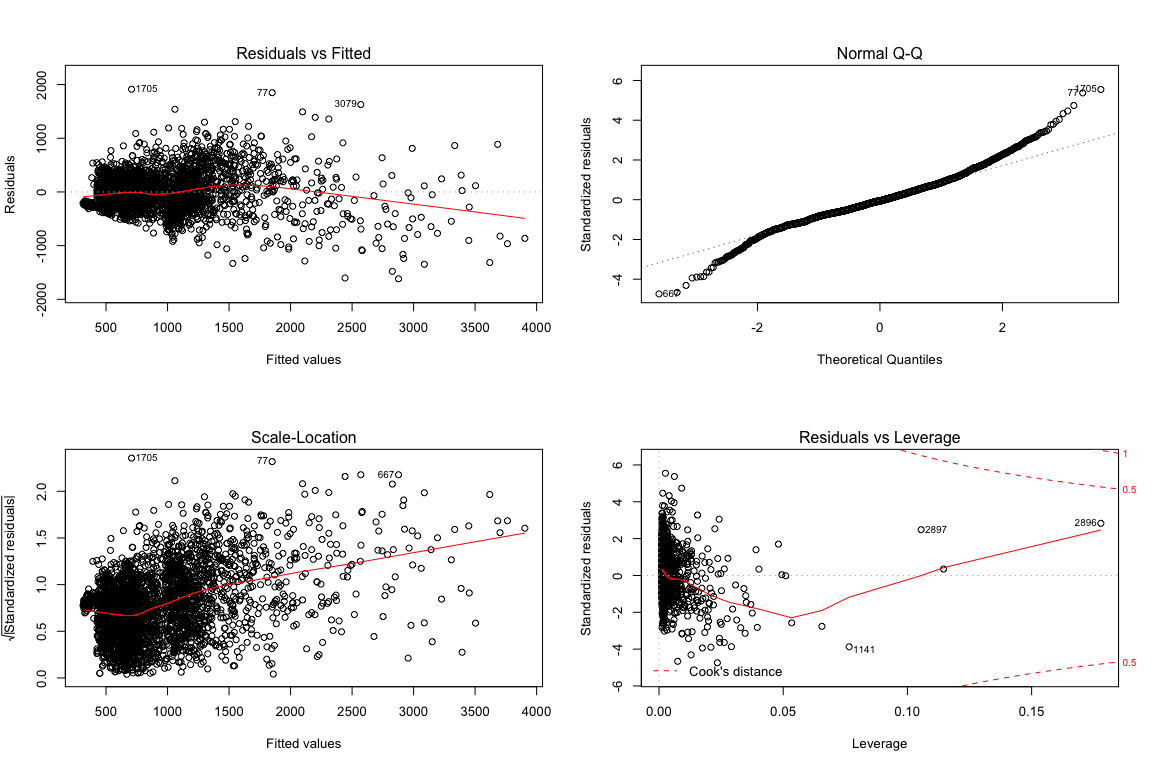
#Create Data.frame for important comparisons  
all\_val <- data.frame(R2 = 1:7, AIC = 1:7)  
  
# Build initial model, ht ~ dbh \* genus  
lm\_init <- lm(Woody\_Height\_cm ~ Woody\_DBH\_cm \* Genus, data = ripdata\_ss)  
anova(lm\_init)

## Analysis of Variance Table  
##   
## Response: Woody\_Height\_cm  
## Df Sum Sq Mean Sq F value Pr(>F)   
## Woody\_DBH\_cm 1 617404685 617404685 5181.224 < 2.2e-16 \*\*\*  
## Genus 4 97049572 24262393 203.609 < 2.2e-16 \*\*\*  
## Woody\_DBH\_cm:Genus 4 9079475 2269869 19.049 1.697e-15 \*\*\*  
## Residuals 3220 383701442 119162   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

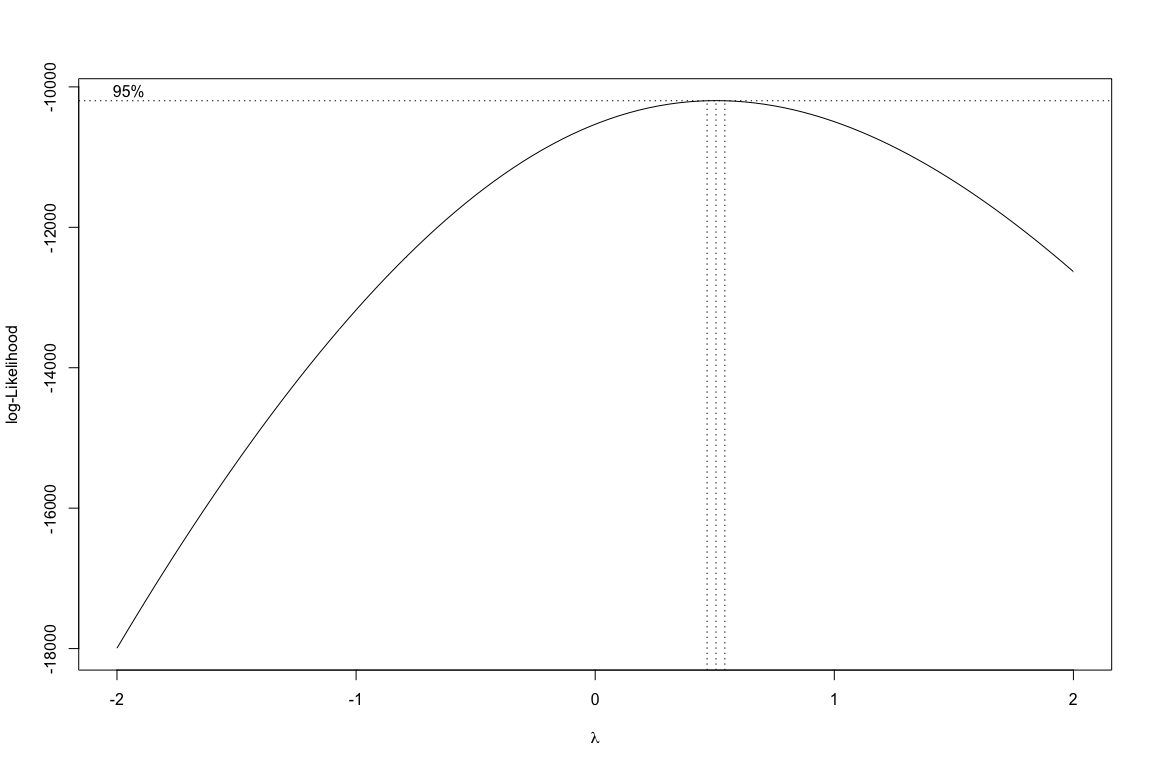
summary(lm\_init)

##   
## Call:  
## lm(formula = Woody\_Height\_cm ~ Woody\_DBH\_cm \* Genus, data = ripdata\_ss)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1617.47 -211.35 -22.81 194.36 1912.49   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 302.180 25.330 11.930 < 2e-16 \*\*\*  
## Woody\_DBH\_cm 30.601 1.486 20.592 < 2e-16 \*\*\*  
## GenusFraxinus -17.743 35.311 -0.502 0.615368   
## GenusPopulus 587.841 30.754 19.114 < 2e-16 \*\*\*  
## GenusQuercus 111.444 31.995 3.483 0.000502 \*\*\*  
## GenusSalix 189.152 33.006 5.731 1.09e-08 \*\*\*  
## Woody\_DBH\_cm:GenusFraxinus -2.206 2.296 -0.961 0.336619   
## Woody\_DBH\_cm:GenusPopulus -11.189 1.568 -7.136 1.18e-12 \*\*\*  
## Woody\_DBH\_cm:GenusQuercus -8.324 1.601 -5.200 2.12e-07 \*\*\*  
## Woody\_DBH\_cm:GenusSalix -5.858 2.156 -2.718 0.006612 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 345.2 on 3220 degrees of freedom  
## Multiple R-squared: 0.6535, Adjusted R-squared: 0.6525   
## F-statistic: 674.7 on 9 and 3220 DF, p-value: < 2.2e-16

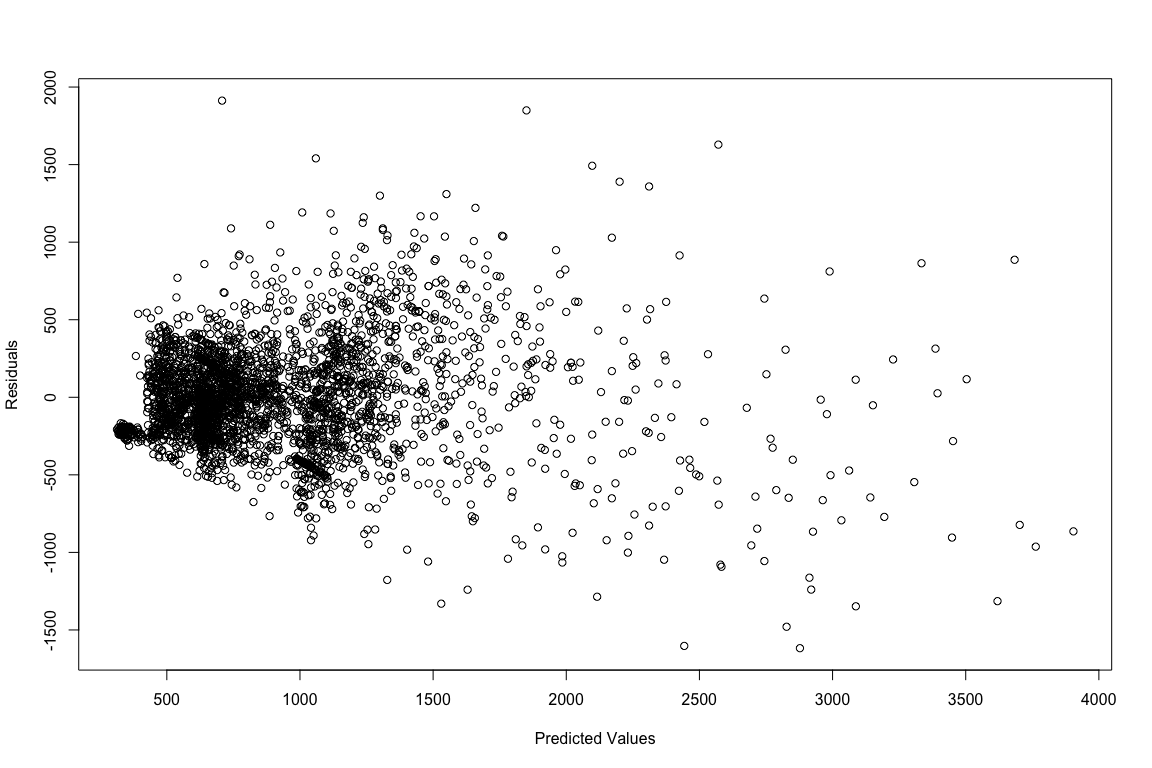
#Notes: Fraxinus does not have a significant p-value  
  
#I am basing my models off of AIC and r.squared: high r.squared and low AIC  
all\_val[1,"R2"] <- summary(lm\_init)$r.squared  
all\_val[1,"AIC"] <- AIC(lm\_init)  
  
  
  
par(mfrow = c(2,2))  
plot(lm\_init)



par(mfrow = c(1,1))  
#The qq plot indicates that normality has not been achieved, and residuals vs fitted has a funnel-esque shape, thus some heteroscedasticity that can be corrected  
  
boxcox(lm\_init)



plot(residuals(lm\_init) ~ predict(lm\_init), xlab = "Predicted Values", ylab = "Residuals")



shapiro.test(residuals(lm\_init)) #shapiro-wilk test for normality of residuals

##   
## Shapiro-Wilk normality test  
##   
## data: residuals(lm\_init)  
## W = 0.97505, p-value < 2.2e-16

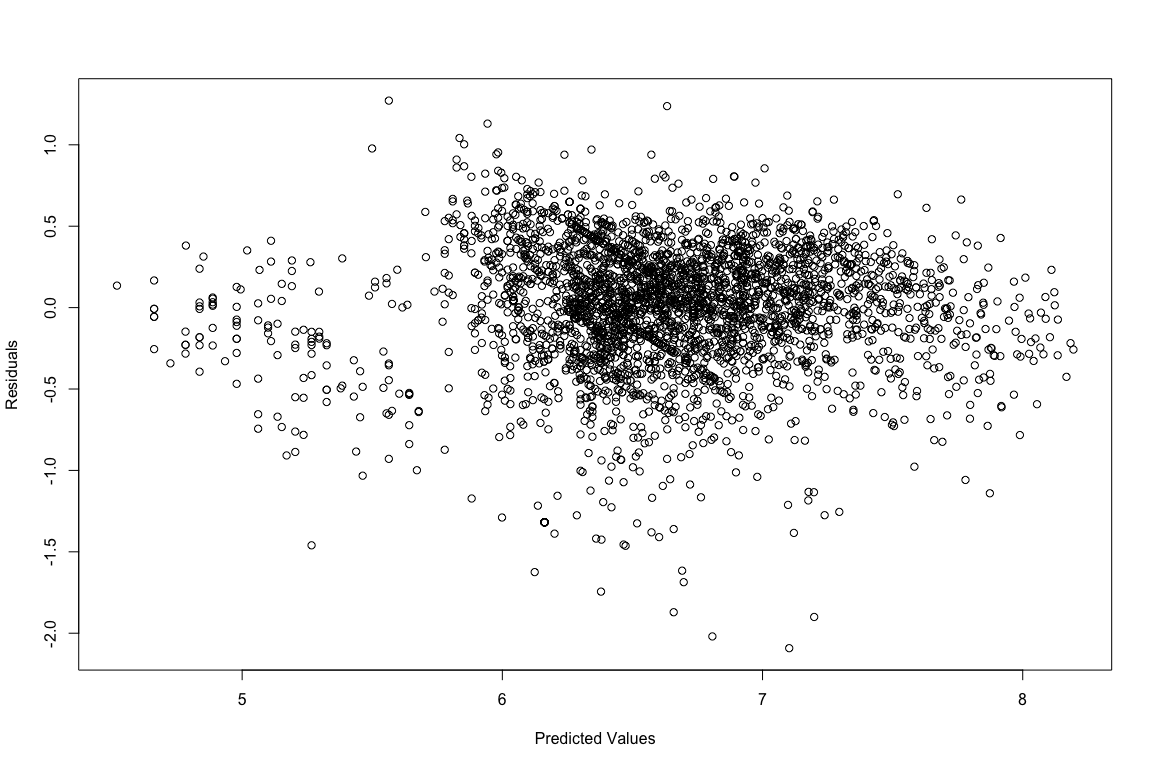
#p-value < 2.2e-16: not normally distributed. There's a funnel shape to the residuals.   
  
lm\_init\_log <- lm(log\_ht ~ log\_DBH \* Genus, data = ripdata\_ss)  
anova(lm\_init\_log)

## Analysis of Variance Table  
##   
## Response: log\_ht  
## Df Sum Sq Mean Sq F value Pr(>F)   
## log\_DBH 1 852.48 852.48 5691.065 < 2.2e-16 \*\*\*  
## Genus 4 95.41 23.85 159.243 < 2.2e-16 \*\*\*  
## log\_DBH:Genus 4 19.50 4.88 32.551 < 2.2e-16 \*\*\*  
## Residuals 3220 482.33 0.15   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

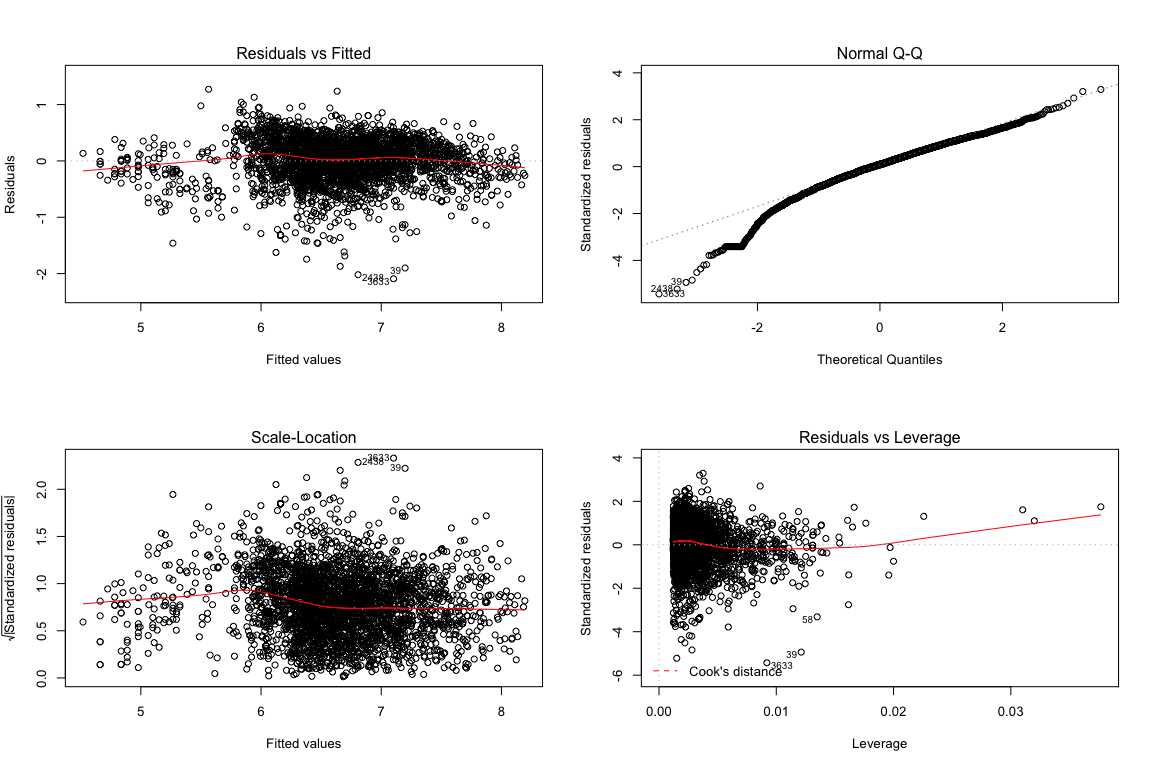
summary(lm\_init\_log)

##   
## Call:  
## lm(formula = log\_ht ~ log\_DBH \* Genus, data = ripdata\_ss)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.09189 -0.20116 0.03404 0.26110 1.27140   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 5.02988 0.07664 65.628 < 2e-16 \*\*\*  
## log\_DBH 0.58103 0.03132 18.549 < 2e-16 \*\*\*  
## GenusFraxinus -0.51093 0.09040 -5.652 1.72e-08 \*\*\*  
## GenusPopulus 0.55644 0.09380 5.932 3.31e-09 \*\*\*  
## GenusQuercus -0.17859 0.09135 -1.955 0.050674 .   
## GenusSalix 0.51754 0.09863 5.247 1.64e-07 \*\*\*  
## log\_DBH:GenusFraxinus 0.20162 0.03809 5.293 1.28e-07 \*\*\*  
## log\_DBH:GenusPopulus -0.05903 0.03612 -1.634 0.102306   
## log\_DBH:GenusQuercus 0.05942 0.03550 1.674 0.094255 .   
## log\_DBH:GenusSalix -0.13930 0.04134 -3.369 0.000763 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.387 on 3220 degrees of freedom  
## Multiple R-squared: 0.6673, Adjusted R-squared: 0.6664   
## F-statistic: 717.6 on 9 and 3220 DF, p-value: < 2.2e-16

plot(residuals(lm\_init\_log) ~ predict(lm\_init\_log), xlab = "Predicted Values", ylab = "Residuals")



#The residual plot is much less funnel shaped.  
  
par(mfrow = c(2,2))  
plot(lm\_init\_log)



par(mfrow = c(1,1))  
  
all\_val[2,"R2"] <- summary(lm\_init\_log)$r.squared  
all\_val[2,"AIC"] <- AIC(lm\_init\_log)  
  
al\_lm1 <- lm(log\_ht ~ log\_DBH \* Genus + tavg\_Aug, data = ripdata\_ss)  
  
anova(al\_lm1)

## Analysis of Variance Table  
##   
## Response: log\_ht  
## Df Sum Sq Mean Sq F value Pr(>F)   
## log\_DBH 1 852.48 852.48 5696.3695 <2e-16 \*\*\*  
## Genus 4 95.41 23.85 159.3909 <2e-16 \*\*\*  
## tavg\_Aug 1 0.29 0.29 1.9078 0.1673   
## log\_DBH:Genus 4 19.82 4.95 33.1047 <2e-16 \*\*\*  
## Residuals 3219 481.73 0.15   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

all\_val[3,"R2"] <- summary(al\_lm1)$r.squared  
all\_val[3,"AIC"] <- AIC(al\_lm1)  
  
al\_lm2 <- lm(log\_ht ~ log\_DBH \* Genus + elevation, data = ripdata\_ss)  
  
anova(al\_lm2)

## Analysis of Variance Table  
##   
## Response: log\_ht  
## Df Sum Sq Mean Sq F value Pr(>F)   
## log\_DBH 1 852.48 852.48 5744.867 < 2.2e-16 \*\*\*  
## Genus 4 95.41 23.85 160.748 < 2.2e-16 \*\*\*  
## elevation 1 4.54 4.54 30.578 3.463e-08 \*\*\*  
## log\_DBH:Genus 4 19.63 4.91 33.074 < 2.2e-16 \*\*\*  
## Residuals 3219 477.67 0.15   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

all\_val[4,"R2"] <- summary(al\_lm2)$r.squared  
all\_val[4,"AIC"] <- AIC(al\_lm2)  
  
al\_lm3 <- lm(log\_ht ~ log\_DBH \* Genus + Latitude, data = ripdata\_ss)  
  
anova(al\_lm3)

## Analysis of Variance Table  
##   
## Response: log\_ht  
## Df Sum Sq Mean Sq F value Pr(>F)   
## log\_DBH 1 852.48 852.48 5718.578 < 2.2e-16 \*\*\*  
## Genus 4 95.41 23.85 160.012 < 2.2e-16 \*\*\*  
## Latitude 1 2.37 2.37 15.891 6.859e-05 \*\*\*  
## log\_DBH:Genus 4 19.60 4.90 32.877 < 2.2e-16 \*\*\*  
## Residuals 3219 479.86 0.15   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

all\_val[5,"R2"] <- summary(al\_lm3)$r.squared  
all\_val[5,"AIC"] <- AIC(al\_lm3)  
  
al\_lm4 <- lm(log\_ht ~ log\_DBH \* Genus + precp\_Aug, data = ripdata\_ss)  
  
anova(al\_lm4)

## Analysis of Variance Table  
##   
## Response: log\_ht  
## Df Sum Sq Mean Sq F value Pr(>F)   
## log\_DBH 1 852.48 852.48 5736.109 < 2.2e-16 \*\*\*  
## Genus 4 95.41 23.85 160.503 < 2.2e-16 \*\*\*  
## precp\_Aug 1 3.93 3.93 26.462 2.849e-07 \*\*\*  
## log\_DBH:Genus 4 19.51 4.88 32.815 < 2.2e-16 \*\*\*  
## Residuals 3219 478.39 0.15   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

all\_val[6,"R2"] <- summary(al\_lm4)$r.squared  
all\_val[6,"AIC"] <- AIC(al\_lm4)  
  
rownames(all\_val) <- c("Initial", "Log\_Initial", "tavg\_Aug", "Elevation", "Latitude", "precp\_Aug", "All\_values")  
  
#Multiple parameters may increase model fit  
al\_lm5 <- lm(log\_ht ~ log\_DBH \* Genus + precp\_Aug + Latitude, data = ripdata\_ss)  
  
anova(al\_lm5)

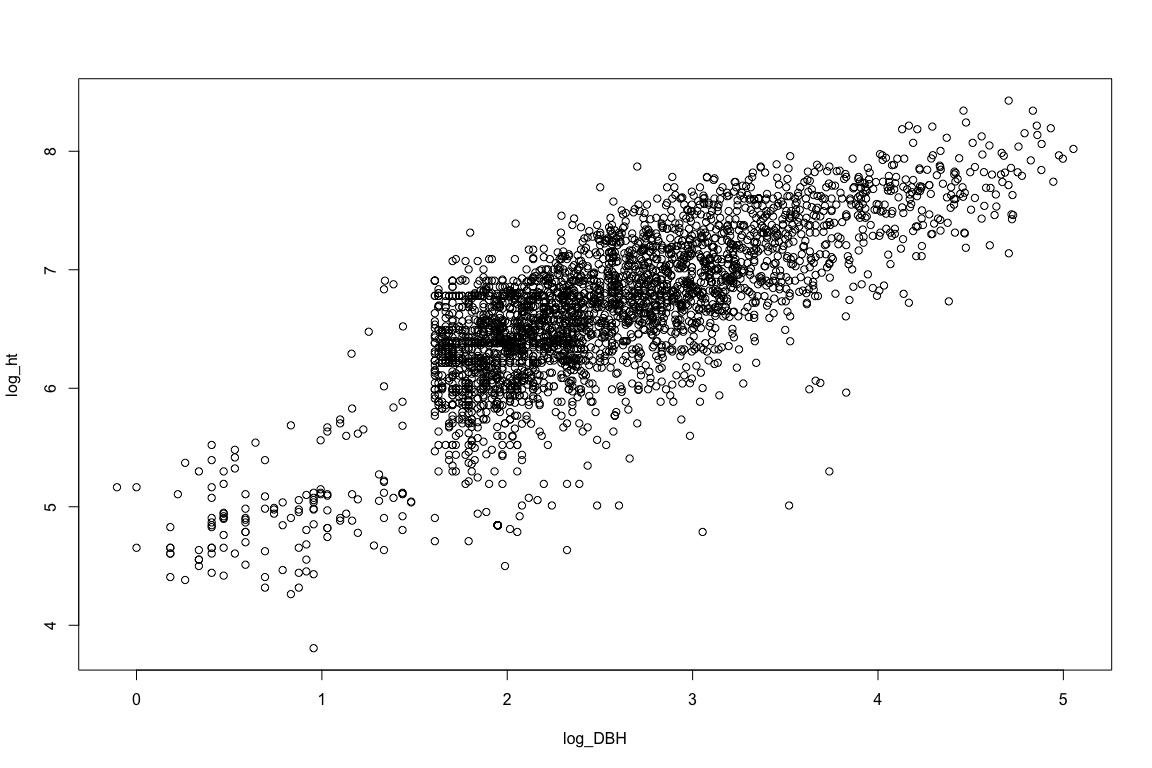
## Analysis of Variance Table  
##   
## Response: log\_ht  
## Df Sum Sq Mean Sq F value Pr(>F)   
## log\_DBH 1 852.48 852.48 5748.2517 < 2.2e-16 \*\*\*  
## Genus 4 95.41 23.85 160.8427 < 2.2e-16 \*\*\*  
## precp\_Aug 1 3.93 3.93 26.5176 2.768e-07 \*\*\*  
## Latitude 1 1.41 1.41 9.5151 0.002055 \*\*   
## log\_DBH:Genus 4 19.25 4.81 32.4589 < 2.2e-16 \*\*\*  
## Residuals 3218 477.24 0.15   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

all\_val[7,"R2"] <- summary(al\_lm5)$r.squared  
all\_val[7,"AIC"] <- AIC(al\_lm5)  
  
all\_val

## R2 AIC  
## Initial 0.6534599 46931.338  
## Log\_Initial 0.6672948 3046.153  
## tavg\_Aug 0.6677079 3044.141  
## Elevation 0.6705130 3016.758  
## Latitude 0.6689983 3031.573  
## precp\_Aug 0.6700100 3021.685  
## All\_values 0.6708094 3015.852

#Including multiple parameters increases the explained variance, but not by much. It appears that including elevation as a parameter would be sufficient. Elevation would make sense as well: elevation can be a proxy for climate. A simple search for "elevation gradient" on Google Scholar will turn up hundreds of papers regarding the geomorphology of plant species that differs based on elevation.

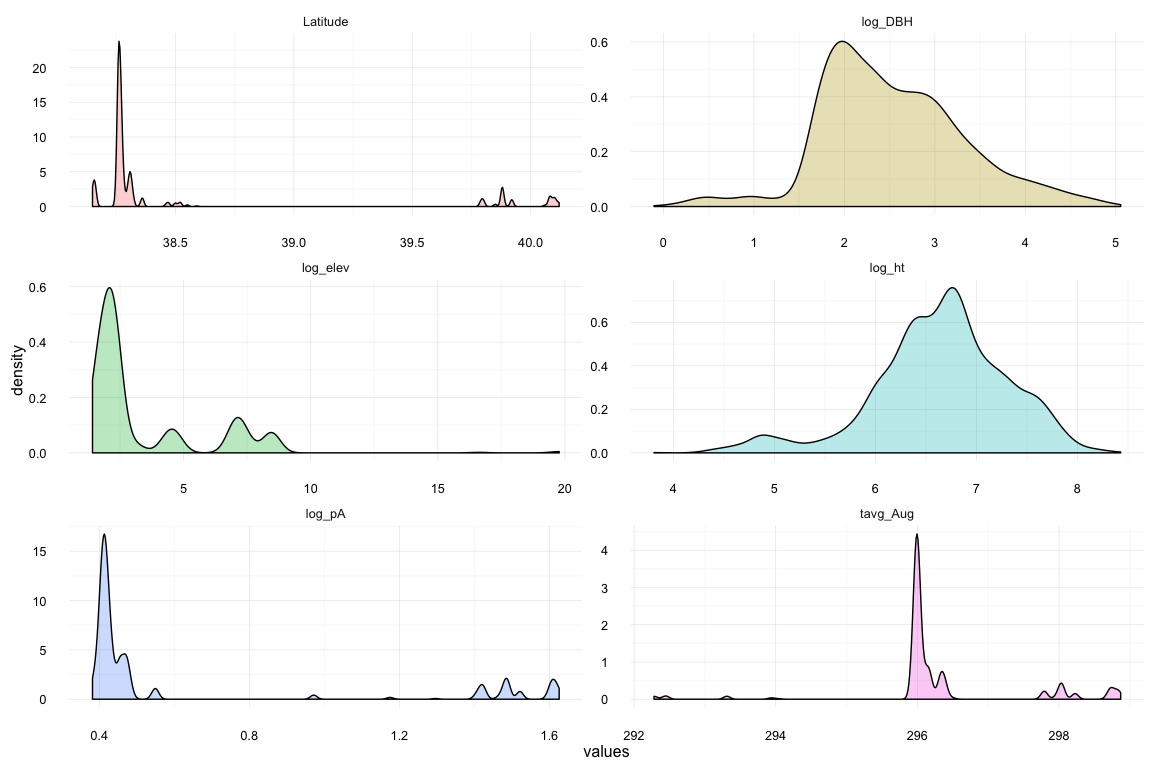
new\_data <- read.csv("~/Desktop/General/Spring 2016/ES 207/Labs/Projects/Lab-1/Data/new\_data.csv", stringsAsFactors=FALSE)  
  
with(ripdata\_ss,plot(log\_DBH, log\_ht))  
  
lat2 <- new\_data$Latitude  
lon2 <- new\_data$Longitude  
xy2 <- cbind(lon2, lat2)  
new\_data$elevation <- extract(r1, xy2)  
new\_data$log\_DBH <- log(new\_data$Woody\_DBH\_cm)  
new\_data$log\_ht <- predict(al\_lm2, newdata= new\_data)  
new\_data$Woody\_ht\_cm <- exp(new\_data$log\_ht)  
lines(new\_data$Woody\_ht\_cm, col = "blue", lwd = 3)



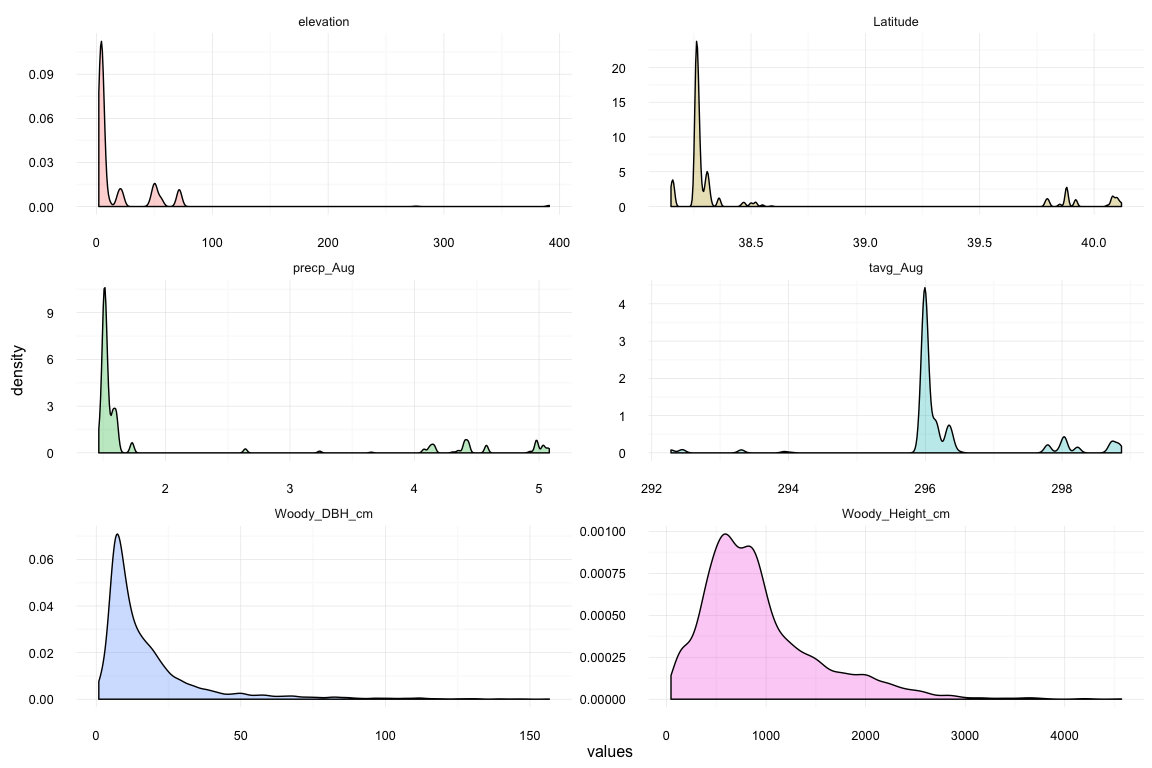
#This doesn't seem correct, but I'll go with it, for now.  
  
agtv <- 705 \* (.0000334750\*(new\_data$Woody\_DBH\_cm ^2.33631) \* (new\_data$Woody\_ht\_cm ^.74872))  
  
#95% confidence intervals can be obtained via the "levels" option of the predict function. By generating those values, we could use the above formula to give a range of values from the 95% confidence interval.

Results: Our linear model required the log transformation of DBH and height to generate a normal distruibution of each. We did not transform other variables because we could not discuver an adequate transformation. All attempted transformations did not result in normal distributions:

tru



nw

 We determined that lm(log\_ht ~ log\_DBH \* Genus + elevation) resulted in the most accurate model, though it should be furthur refined.

all\_val

## R2 AIC  
## Initial 0.6534599 46931.338  
## Log\_Initial 0.6672948 3046.153  
## tavg\_Aug 0.6677079 3044.141  
## Elevation 0.6705130 3016.758  
## Latitude 0.6689983 3031.573  
## precp\_Aug 0.6700100 3021.685  
## All\_values 0.6708094 3015.852

Based on the linear model using elevation, genus, and the interaction of genus and diameter at breast height, we predict the total carbon stock of trees at the new site to be 7.596005510^{5} Mg per hectare.

Discussion:  
Our chosen parameter, elevation, is a fitting varible to complete our model. Not only did a comparison of AIC (3016.758) and r.squared (.6705130) show that this was the best combination of any parameter, but the literature supports the use of elevation as a proxy for climatic differences (Körner 2007, Malhi 2010, Whittaker 1975). A simple search for "elevation gradient" on Google Scholar will turn up hundreds of papers regarding the geomorphology of plant species that differs based on elevation.  
However, our model only explained 67% of the variation. This is not enough to accurately generate predictions. It is possible that because a zero intercept was not forced or included that our estimates underestimate carbon stocks.

Limitations: We do not include longitude in our analysis because elevation's in California change over east-west rather than north-south

Citations:

H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2009.

H. Wickham. Reshaping data with the reshape package. Journal of Statistical Software, 21(12), 2007.

Körner, Christian. "The use of ‘altitude’in ecological research." Trends in ecology & evolution 22.11 (2007): 569-574.

Malhi, Yadvinder, et al. "Introduction: elevation gradients in the tropics: laboratories for ecosystem ecology and global change research." Global Change Biology 16.12 (2010): 3171-3175.

R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

Richard M. Heiberger (2016). HH: Statistical Analysis and Data Display: Heiberger and Holland. R package version 3.1-25. URL <http://CRAN.R-project.org/package=HH>

Roger Bivand, Tim Keitt and Barry Rowlingson (2015). rgdal: Bindings for the Geospatial Data Abstraction Library. R package version 1.0-4. <https://CRAN.R-project.org/package=rgdal>

Robert J. Hijmans (2015). raster: Geographic Data Analysis and Modeling. R package version 2.5-2. <https://CRAN.R-project.org/package=raster>

Venables, W. N. & Ripley, B. D. (2002) Modern Applied Statistics with S. Fourth Edition. Springer, New York. ISBN 0-387-95457-0

Whittaker, Robert H., and William A. Niering. "Vegetation of the Santa Catalina Mountains, Arizona. V. Biomass, production, and diversity along the elevation gradient." Ecology (1975): 771-790.