Joseph Crockett

ES 207: Environmental Data Analysis

March 1, 2016

**Homework Assignment 4**

**Objective Statement:**

In order to reasonably estimate standing carbon, we will further refine our linear model between Diameter at Breast Height and Tree height, factored by either genus or project site.

**Methods:**

We determined the mean, standard deviation, and count, for the five subsetted genera by plot site. Plots at which more than one measurement was taken were retained. Of these, we randomly sampled 6 plot sites for each of the four project areas. The coefficient of variation was calculated, and an ANOVA test and a TukeyHSD test were run to determine whether project area was a determining factor for tree height. We developed two linear models using genus as single terms and as an interaction variable describing Height (cm) ~ DBH (cm). The models were coerced with a zero intercept to reflect seedling height and DBH at germination.

**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 were identified and removed as well.

**Code:**

rd\_lab4$projplot <- as.factor( paste(rd\_lab4$ProjCode,rd\_lab4$Plot.Name))  
  
#Generating mean, sd, and count of 5 most frequent genera at each plot.  
rd\_sum <- data.frame(cbind(tapply(rd\_lab4$Woody\_Height\_cm, rd\_lab4$projplot, mean), tapply(rd\_lab4$Woody\_Height\_cm,rd\_lab4$projplot,sd),tapply(rd\_lab4$Woody\_Height\_cm,rd\_lab4$projplot,length)))  
  
colnames(rd\_sum) <- c("htcmmn","htcmsd","plot.n")  
  
rd\_sum$projplot <- as.factor(rownames(rd\_sum))  
#needs greater than 1 measurement  
rd\_sum <- rd\_sum[rd\_sum$plot.n >1 ,]  
#new column for labels  
rd\_sum$proj <- as.factor(substr(rd\_sum$projplot,1,5))  
  
projlevels <- levels(rd\_sum$proj)  
  
for (p in 1:length(projlevels)){  
 print(summary(rd\_sum[rd\_sum$proj == projlevels[p],]))  
}

## htcmmn htcmsd plot.n projplot   
## Min. : 137.1 Min. : 42.43 Min. : 2.00 CORP 1 : 1   
## 1st Qu.: 768.7 1st Qu.: 247.74 1st Qu.: 10.00 CORP 2 : 1   
## Median :1000.0 Median : 412.96 Median : 18.00 CORP 3 : 1   
## Mean :1035.2 Mean : 439.64 Mean : 25.15 CORP 4 : 1   
## 3rd Qu.:1229.1 3rd Qu.: 615.85 3rd Qu.: 32.00 CORP 5 : 1   
## Max. :3330.0 Max. :1031.60 Max. :111.00 CORP 6 : 1   
## (Other):83   
## proj   
## CORP :89   
## HOWY : 0   
## NASO : 0   
## SRRB : 0   
##   
##   
##   
## htcmmn htcmsd plot.n projplot  
## Min. : 421.3 Min. : 68.18 Min. : 7.00 HOWY RIP01:1   
## 1st Qu.: 505.6 1st Qu.: 141.83 1st Qu.:11.00 HOWY RIP02:1   
## Median : 883.0 Median : 218.21 Median :27.00 HOWY RIP03:1   
## Mean : 964.1 Mean : 409.68 Mean :38.86 HOWY RIP04:1   
## 3rd Qu.:1292.6 3rd Qu.: 528.12 3rd Qu.:62.00 HOWY RIP05:1   
## Max. :1848.3 Max. :1241.50 Max. :92.00 HOWY RIP06:1   
## (Other) :1   
## proj   
## CORP :0   
## HOWY :7   
## NASO :0   
## SRRB :0   
##   
##   
##   
## htcmmn htcmsd plot.n projplot  
## Min. : 671.0 Min. :171.6 Min. : 4.00 NASO Crp2013\_509:1   
## 1st Qu.: 745.4 1st Qu.:472.4 1st Qu.: 6.75 NASO S02013\_200 :1   
## Median : 814.6 Median :556.3 Median :17.50 NASO S02013\_202 :1   
## Mean : 977.8 Mean :538.2 Mean :24.00 NASO S02013\_203 :1   
## 3rd Qu.:1016.4 3rd Qu.:649.9 3rd Qu.:29.25 NASO S02013\_207 :1   
## Max. :1888.3 Max. :819.5 Max. :81.00 NASO S02013\_211 :1   
## (Other) :2   
## proj   
## CORP :0   
## HOWY :0   
## NASO :8   
## SRRB :0   
##   
##   
##   
## htcmmn htcmsd plot.n projplot   
## Min. : 443.2 Min. : 57.79 Min. : 2.0 SRRB EW2013\_100: 1   
## 1st Qu.: 664.8 1st Qu.: 246.65 1st Qu.: 4.0 SRRB EW2013\_101: 1   
## Median : 977.0 Median : 572.42 Median : 7.0 SRRB EW2013\_102: 1   
## Mean :1065.4 Mean : 610.42 Mean :11.8 SRRB EW2013\_103: 1   
## 3rd Qu.:1256.6 3rd Qu.: 882.65 3rd Qu.:17.0 SRRB EW2013\_106: 1   
## Max. :2532.9 Max. :1668.77 Max. :59.0 SRRB EW2013\_110: 1   
## (Other) :38   
## proj   
## CORP : 0   
## HOWY : 0   
## NASO : 0   
## SRRB :44   
##   
##   
##

lapply(projlevels, function(x) summary(rd\_sum[rd\_sum$proj == x,]))

## [[1]]  
## htcmmn htcmsd plot.n projplot   
## Min. : 137.1 Min. : 42.43 Min. : 2.00 CORP 1 : 1   
## 1st Qu.: 768.7 1st Qu.: 247.74 1st Qu.: 10.00 CORP 2 : 1   
## Median :1000.0 Median : 412.96 Median : 18.00 CORP 3 : 1   
## Mean :1035.2 Mean : 439.64 Mean : 25.15 CORP 4 : 1   
## 3rd Qu.:1229.1 3rd Qu.: 615.85 3rd Qu.: 32.00 CORP 5 : 1   
## Max. :3330.0 Max. :1031.60 Max. :111.00 CORP 6 : 1   
## (Other):83   
## proj   
## CORP :89   
## HOWY : 0   
## NASO : 0   
## SRRB : 0   
##   
##   
##   
##   
## [[2]]  
## htcmmn htcmsd plot.n projplot  
## Min. : 421.3 Min. : 68.18 Min. : 7.00 HOWY RIP01:1   
## 1st Qu.: 505.6 1st Qu.: 141.83 1st Qu.:11.00 HOWY RIP02:1   
## Median : 883.0 Median : 218.21 Median :27.00 HOWY RIP03:1   
## Mean : 964.1 Mean : 409.68 Mean :38.86 HOWY RIP04:1   
## 3rd Qu.:1292.6 3rd Qu.: 528.12 3rd Qu.:62.00 HOWY RIP05:1   
## Max. :1848.3 Max. :1241.50 Max. :92.00 HOWY RIP06:1   
## (Other) :1   
## proj   
## CORP :0   
## HOWY :7   
## NASO :0   
## SRRB :0   
##   
##   
##   
##   
## [[3]]  
## htcmmn htcmsd plot.n projplot  
## Min. : 671.0 Min. :171.6 Min. : 4.00 NASO Crp2013\_509:1   
## 1st Qu.: 745.4 1st Qu.:472.4 1st Qu.: 6.75 NASO S02013\_200 :1   
## Median : 814.6 Median :556.3 Median :17.50 NASO S02013\_202 :1   
## Mean : 977.8 Mean :538.2 Mean :24.00 NASO S02013\_203 :1   
## 3rd Qu.:1016.4 3rd Qu.:649.9 3rd Qu.:29.25 NASO S02013\_207 :1   
## Max. :1888.3 Max. :819.5 Max. :81.00 NASO S02013\_211 :1   
## (Other) :2   
## proj   
## CORP :0   
## HOWY :0   
## NASO :8   
## SRRB :0   
##   
##   
##   
##   
## [[4]]  
## htcmmn htcmsd plot.n projplot   
## Min. : 443.2 Min. : 57.79 Min. : 2.0 SRRB EW2013\_100: 1   
## 1st Qu.: 664.8 1st Qu.: 246.65 1st Qu.: 4.0 SRRB EW2013\_101: 1   
## Median : 977.0 Median : 572.42 Median : 7.0 SRRB EW2013\_102: 1   
## Mean :1065.4 Mean : 610.42 Mean :11.8 SRRB EW2013\_103: 1   
## 3rd Qu.:1256.6 3rd Qu.: 882.65 3rd Qu.:17.0 SRRB EW2013\_106: 1   
## Max. :2532.9 Max. :1668.77 Max. :59.0 SRRB EW2013\_110: 1   
## (Other) :38   
## proj   
## CORP : 0   
## HOWY : 0   
## NASO : 0   
## SRRB :44   
##   
##   
##

#randomly selecting 6 sample plot summaries from each project site  
nsamples <- 6  
ripres <-lapply(projlevels, function(x) rd\_sum[which(rd\_sum$proj == x),][sample(nrow(rd\_sum[which(rd\_sum$proj ==x),]),nsamples),])  
  
#do.call, similar to sapply, but not simple. only works on lists (hence lapply). Combining by each value (1-4) by rows.  
ripsample <- do.call(rbind,ripres)  
summary(ripsample$proj)

## CORP HOWY NASO SRRB   
## 6 6 6 6

#Find Coefficient of variation  
ripsample$cv <- with(ripsample, htcmsd/htcmmn)  
  
#one-way ANOVA   
rip.proj.cv.aov <- aov(cv~proj, data = ripsample)  
summary(rip.proj.cv.aov)

## Df Sum Sq Mean Sq F value Pr(>F)   
## proj 3 0.2483 0.08277 2.692 0.0736 .  
## Residuals 20 0.6149 0.03074   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

#vs  
summary.lm(rip.proj.cv.aov)

##   
## Call:  
## aov(formula = cv ~ proj, data = ripsample)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.22751 -0.14853 0.00352 0.12997 0.34894   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.51954 0.07158 7.258 5.07e-07 \*\*\*  
## projHOWY -0.19370 0.10123 -1.913 0.0701 .   
## projNASO 0.06827 0.10123 0.674 0.5078   
## projSRRB 0.03522 0.10123 0.348 0.7315   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.1753 on 20 degrees of freedom  
## Multiple R-squared: 0.2877, Adjusted R-squared: 0.1808   
## F-statistic: 2.692 on 3 and 20 DF, p-value: 0.07365

#summary.lm returns a linear regression summary of cv ~ proj, including expanded coefficients, std. error, r2, residuals, p and t values. Summary of the anova returns proj as a single variable, the residuals, sum of squares, mean squares, f value and degrees of freedom.   
  
rip.aov.hsd <- TukeyHSD(rip.proj.cv.aov)  
rip.aov.hsd

## Tukey multiple comparisons of means  
## 95% family-wise confidence level  
##   
## Fit: aov(formula = cv ~ proj, data = ripsample)  
##   
## $proj  
## diff lwr upr p adj  
## HOWY -CORP -0.19369756 -0.47704098 0.08964586 0.2541655  
## NASO -CORP 0.06827416 -0.21506927 0.35161758 0.9055302  
## SRRB -CORP 0.03522481 -0.24811861 0.31856823 0.9851007  
## NASO -HOWY 0.26197172 -0.02137170 0.54531514 0.0763170  
## SRRB -HOWY 0.22892237 -0.05442105 0.51226579 0.1411081  
## SRRB -NASO -0.03304935 -0.31639277 0.25029407 0.9876183

#Based on the adjusted p value (last column), there is no signifant difference between sites. From the one-way ANOVA, the predictors do not have a significant impact on the response.

#Two models: First, with single terms (hcm ~ dbhcm), second with interaction (hcm ~ dbhcm \* genus). Looking first at hcm ~ dbhcm by Genus group.  
# + vs \*: + calculates main effects only while \* estimates interactions between factors.  
#In addition, forcing a zero intercept.  
incpt <- 0.0  
  
  
g1 <- ggplot(data = rd\_lab4,aes(x = Woody\_DBH\_cm, y = Woody\_Height\_cm, color = Genus)) + geom\_point() + theme\_minimal() + labs(title = "", x = "DBH, cm", y = "Height, cm") + theme(legend.position = c(0,1), legend.justification = c(0,1))  
  
  
#Model 1, simple global model  
glb\_lm <- aov(I(Woody\_Height\_cm - incpt) ~ 0 + Woody\_DBH\_cm, data = rd\_lab4)  
  
  
#Model 2, global model with single terms.   
glb\_lm2 <- aov(I(Woody\_Height\_cm - incpt) ~ (0 + Woody\_DBH\_cm) + Genus, data = rd\_lab4)  
  
  
#Model 3, global model with interaction terms.  
glb\_lm3 <- aov(I(Woody\_Height\_cm - incpt) ~ (0 + Woody\_DBH\_cm) \* Genus, data = rd\_lab4)  
  
  
g2 <- ancovaplot(Woody\_Height\_cm ~ Woody\_DBH\_cm + Genus, data = rd\_lab4, title = "Woody\_Height\_cm ~ Woody\_DBH\_cm + Genus")  
g3 <- ancovaplot(Woody\_Height\_cm ~ Woody\_DBH\_cm \* Genus, data = rd\_lab4, title = "Woody\_Height\_cm ~ Woody\_DBH\_cm \* Genus")  
  
summary(glb\_lm)

## Df Sum Sq Mean Sq F value Pr(>F)   
## Woody\_DBH\_cm 1 2.968e+09 2.968e+09 10407 <2e-16 \*\*\*  
## Residuals 3229 9.208e+08 2.852e+05   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

summary.lm(glb\_lm)

##   
## Call:  
## aov(formula = I(Woody\_Height\_cm - incpt) ~ 0 + Woody\_DBH\_cm,   
## data = rd\_lab4)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2955.36 80.76 272.98 499.64 2062.54   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## Woody\_DBH\_cm 37.4137 0.3668 102 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 534 on 3229 degrees of freedom  
## Multiple R-squared: 0.7632, Adjusted R-squared: 0.7631   
## F-statistic: 1.041e+04 on 1 and 3229 DF, p-value: < 2.2e-16

summary(glb\_lm2)

## Df Sum Sq Mean Sq F value Pr(>F)   
## Woody\_DBH\_cm 1 2.968e+09 2.968e+09 24358.7 <2e-16 \*\*\*  
## Genus 5 5.280e+08 1.056e+08 866.8 <2e-16 \*\*\*  
## Residuals 3224 3.928e+08 1.218e+05   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

summary.lm(glb\_lm2)

##   
## Call:  
## aov(formula = I(Woody\_Height\_cm - incpt) ~ (0 + Woody\_DBH\_cm) +   
## Genus, data = rd\_lab4)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1727.98 -220.26 -25.81 192.40 1940.69   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## Woody\_DBH\_cm 21.6924 0.3572 60.74 <2e-16 \*\*\*  
## GenusAcer 419.1678 16.9878 24.68 <2e-16 \*\*\*  
## GenusFraxinus 356.0974 16.5970 21.45 <2e-16 \*\*\*  
## GenusPopulus 832.5859 15.1610 54.92 <2e-16 \*\*\*  
## GenusQuercus 427.9189 15.8281 27.04 <2e-16 \*\*\*  
## GenusSalix 524.9874 13.0360 40.27 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 349 on 3224 degrees of freedom  
## Multiple R-squared: 0.899, Adjusted R-squared: 0.8988   
## F-statistic: 4782 on 6 and 3224 DF, p-value: < 2.2e-16

summary(glb\_lm3)

## Df Sum Sq Mean Sq F value Pr(>F)   
## Woody\_DBH\_cm 1 2.968e+09 2.968e+09 24904.12 < 2e-16 \*\*\*  
## Genus 5 5.280e+08 1.056e+08 886.21 < 2e-16 \*\*\*  
## Woody\_DBH\_cm:Genus 4 9.079e+06 2.270e+06 19.05 1.7e-15 \*\*\*  
## Residuals 3220 3.837e+08 1.192e+05   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

summary.lm(glb\_lm3)

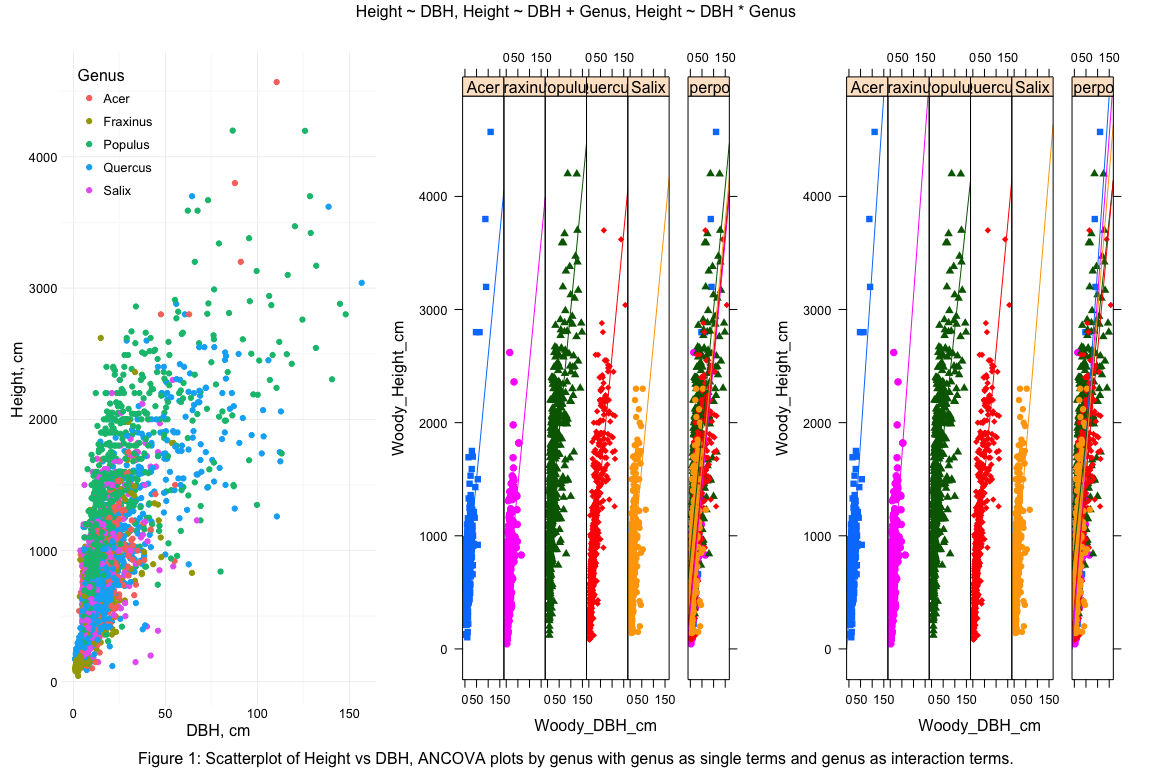
##   
## Call:  
## aov(formula = I(Woody\_Height\_cm - incpt) ~ (0 + Woody\_DBH\_cm) \*   
## Genus, data = rd\_lab4)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1617.47 -211.35 -22.81 194.36 1912.49   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## Woody\_DBH\_cm 30.601 1.486 20.592 < 2e-16 \*\*\*  
## GenusAcer 302.180 25.330 11.930 < 2e-16 \*\*\*  
## GenusFraxinus 284.437 24.602 11.561 < 2e-16 \*\*\*  
## GenusPopulus 890.021 17.442 51.029 < 2e-16 \*\*\*  
## GenusQuercus 413.624 19.548 21.160 < 2e-16 \*\*\*  
## GenusSalix 491.332 21.162 23.218 < 2e-16 \*\*\*  
## Woody\_DBH\_cm:GenusFraxinus -2.206 2.296 -0.961 0.33662   
## 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 0.00661 \*\*   
## ---  
## 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: 0.9013, Adjusted R-squared: 0.901   
## F-statistic: 2941 on 10 and 3220 DF, p-value: < 2.2e-16

#Model 3 explains the most variation of the three models (R2 = .901). The second model of the single terms indicates that the Fraxinus Genus is a significant predictor; however, model 3 indicates that the Fraxinus and interaction term between Fraxinus and DBH is not significant (pr(>|t|) == .6154 and .3366, respectively).

**Results:**

Based upon our random sample of plot by project area, we found tree height does not significantly vary based upon plot within project site (p value = .587, R2 = .09). Our initial model based upon untransformed values of Ht~ DBH (from lab 3) explained approximately 76% of the variance. We increased our R2 in that case by log10 transforming variables. We did not transform values for this report. The third model relating HT ~ DBH with genus as a factor and including interaction terms explained the majority of variance compared to the initial global model and a second model including single terms (R2 = .901, .76, and .89, respectively).

grid.arrange(g1,g2,g3, ncol = 3, top = "Height ~ DBH, Height ~ DBH + Genus, Height ~ DBH \* Genus", bottom = "Figure 1: Scatterplot of Height vs DBH, ANCOVA plots by genus with genus as single terms and genus as interaction terms.")



**Discussion:**

Our finding that project site does not have a significant effect on the relationship between height and DBH suggests that we use either height or DBH to estimate the effects of environmental conditions, rather than modeled biomass. The relationship is dependent upon genus, as fig 1 shows. The interaction with the genus Fraxinus was not significant in the third model, though previous models indicate that it is. Fraxinus height and DBH are limited to lower values than the other genera; this, along with the scattered values at the limits of height and dbh, may sway the Fraxinus model to insignificance.

**Limitations:**

Our subset of 5 genera limits our conclusions, especially as the counts of the 6th most frequent genus was close to the 5th. It is unlikely that we would see site playing a different role on the height~dbh relationship in the other genera; however, a full report should address that.