Tree Analysis

Joseph Crockett

### Homework Assignment 1

#### Objective Statement

Technological advances in LiDAR systems have both increased the accuracy and reduced cost in recent years, however, both system cost and labor required to process data is prohibitive at the biome-scale. Using LiDAR measurements of Sierra Nevada conifers, our objective is to determine, if any, the relationship between crown spread and tree height. A positive correlaton would help to create a model that utilizes less costly remote sensing systems (i.e. Landsat/ NAIP)

#### Methods

### Data

tree\_data <- read.csv("~/Desktop/Documents/Writings/Spring 2016/ES 207/Labs/Lab1/Trees.csv",   
 header = T) #load LiDAR tree measurements into local environment data frame   
  
dim(tree\_data) #Checking formatting of frame.

## [1] 176100 5

str(tree\_data)

## 'data.frame': 176100 obs. of 5 variables:  
## $ OBJECTID: int 1 2 3 4 5 6 7 8 9 10 ...  
## $ x : num 707921 707921 708190 708546 707569 ...  
## $ y : num 4333002 4333002 4333002 4333002 4333002 ...  
## $ z.m. : num 2.01 2.01 12.6 42.29 15.63 ...  
## $ r.m. : num 1.03 1.03 2.11 5.47 2.93 ...

anyDuplicated(tree\_data[, 2:5]) # Check for duplicated entries

## [1] 2

# The formatting seems appropriate, but I will rename column  
# names to avoid conflicts.  
colnames(tree\_data) <- c("OBJECTID", "xval", "yval", "zval",   
 "rval")  
  
save(tree\_data, file = "~/Desktop/tree\_data.rda")

ncol(tree\_data) #running data diagnostics

## [1] 5

nrow(tree\_data)

## [1] 176100

head(tree\_data)

## OBJECTID xval yval zval rval  
## 1 1 707921.2 4333002 2.0061 1.032689  
## 2 2 707921.2 4333002 2.0061 1.032689  
## 3 3 708189.9 4333002 12.5996 2.113910  
## 4 4 708546.1 4333002 42.2903 5.471469  
## 5 5 707569.2 4333002 15.6254 2.930143  
## 6 6 708935.7 4333002 13.4347 2.908795

tail(tree\_data)

## OBJECTID xval yval zval rval  
## 176095 176095 708708.2 4334600 5.2456 2.377610  
## 176096 176096 708753.0 4334600 5.0028 1.589963  
## 176097 176097 708753.0 4334600 5.0028 1.452385  
## 176098 176098 708762.9 4334600 44.8285 2.970038  
## 176099 176099 708817.8 4334600 8.1583 2.971006  
## 176100 176100 708817.8 4334600 8.1583 2.971006

summary(tree\_data) # shows mean, max, min, median, IQR, range, etc.

## OBJECTID xval yval zval   
## Min. : 1 Min. :707062 Min. :4333002 Min. : 2.000   
## 1st Qu.: 44026 1st Qu.:707570 1st Qu.:4333401 1st Qu.: 6.155   
## Median : 88050 Median :707997 Median :4333765 Median :10.901   
## Mean : 88050 Mean :708017 Mean :4333777 Mean :13.910   
## 3rd Qu.:132075 3rd Qu.:708471 3rd Qu.:4334154 3rd Qu.:17.776   
## Max. :176100 Max. :709000 Max. :4334600 Max. :73.927   
## rval   
## Min. : 0.1847   
## 1st Qu.: 1.5021   
## Median : 2.2416   
## Mean : 2.5956   
## 3rd Qu.: 3.3191   
## Max. :12.0044

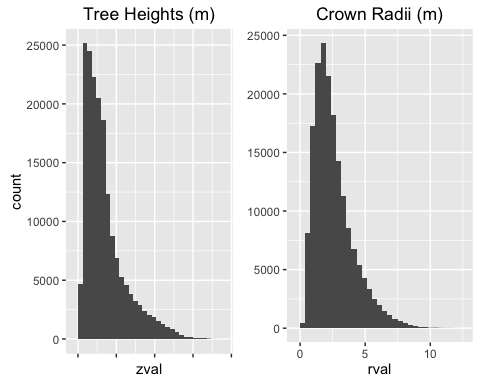
library(ggplot2) #plotting histograms using ggplot2, with a function to plot multiple graphs  
library(gtable)

## Loading required package: grid

library(grid)  
library(gridExtra)  
library(ggExtra)  
  
p1 <- ggplot(tree\_data, aes(x = zval)) + geom\_histogram() + theme(axis.text.x = element\_blank()) +   
 ggtitle("Tree Heights (m)")  
  
p2 <- ggplot(tree\_data, aes(x = rval)) + geom\_histogram() + theme(axis.title.y = element\_blank()) +   
 ggtitle("Crown Radii (m)")  
  
grid.arrange(p1, p2, nrow = 1, ncol = 2)

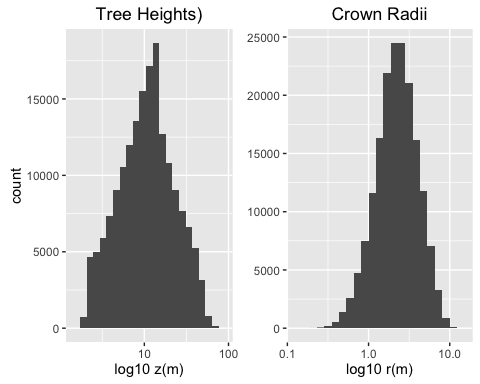
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

 Histograms of tree height and crown radii The data do not appear to meet the assumptions of normality; tree height and crown radii are skewed right.

By log transforming heigth and radii we can better meet normality assumptions.

p3 <- ggplot(tree\_data, aes(x = zval)) + geom\_histogram(bins = 20) +   
 scale\_x\_log10() + labs(title = "Tree Heights)", x = "log10 z(m)",   
 size = 0.5)  
  
p4 <- ggplot(tree\_data, aes(x = rval)) + geom\_histogram(bins = 20) +   
 scale\_x\_log10() + theme(axis.title.y = element\_blank()) +   
 labs(title = "Crown Radii", x = "log10 r(m)", size = 0.5)  
  
grid.arrange(p3, p4, ncol = 2)



# Log transformation gives us normal distribution for tree  
# heights and crown radii.

Pearson Correlations

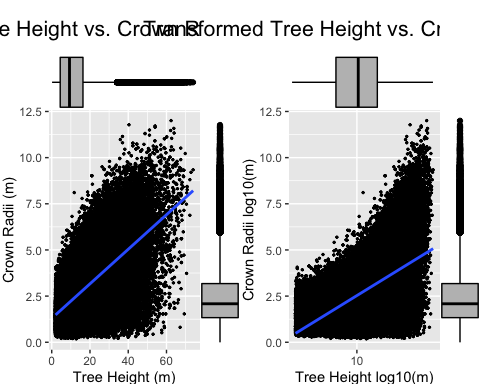
p5 <- ggplot(data = tree\_data, aes(x = zval, y = rval)) + geom\_point(position = "jitter",   
 size = 0.5) + geom\_smooth(method = glm) + labs(title = "Tree Height vs. Crown Radii",   
 x = "Tree Height (m)", y = "Crown Radii (m)")  
  
p6 <- ggMarginal(p5, type = "boxplot")

## Warning: Continuous x aesthetic -- did you forget aes(group=...)?

p7 <- ggplot(data = tree\_data, aes(x = zval, y = rval)) + geom\_point(position = "jitter",   
 size = 0.5) + geom\_smooth(method = glm) + scale\_x\_log10() +   
 labs(title = "Transformed Tree Height vs. Crown Radii", x = "Tree Height log10(m)",   
 y = "Crown Radii log10(m)")  
  
p8 <- ggMarginal(p7, type = "boxplot")

## Warning: Continuous x aesthetic -- did you forget aes(group=...)?

grid.arrange(p6, p8, ncol = 2, nrow = 1)



# Test for correlation between Tree Height and Crown Radii An  
# assumption of Pearson's correlation is that variables must  
# be approximately normally distributed, so log  
# transformation will ensure that this assumption is held.  
  
cor.test(x = log10(tree\_data[, 4]), y = log10(tree\_data[, 5]),   
 na.action = na.omit)

##   
## Pearson's product-moment correlation  
##   
## data: log10(tree\_data[, 4]) and log10(tree\_data[, 5])  
## t = 313.23, df = 176100, p-value < 2.2e-16  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## 0.5951581 0.6011570  
## sample estimates:  
## cor   
## 0.598166

# Without assuming normality, we can use Spearman's rho test  
# to test correlation between non-parametric variables  
  
cor.test(x = tree\_data[, 4], y = tree\_data[, 5], na.action = na.omit,   
 method = "spearman")

## Warning in cor.test.default(x = tree\_data[, 4], y = tree\_data[, 5],  
## na.action = na.omit, : Cannot compute exact p-value with ties

##   
## Spearman's rank correlation rho  
##   
## data: tree\_data[, 4] and tree\_data[, 5]  
## S = 3.4322e+14, p-value < 2.2e-16  
## alternative hypothesis: true rho is not equal to 0  
## sample estimates:  
## rho   
## 0.6229113

Both Spearman's rho and Pearson's product moment correlation show statistical significance (p-value < 2.2e-16 and < 2.2e-16, respectively). We reject the null hypothesis of no correlation between tree height and crown radii. R- values are .623 and .598 respectively.

EDA <- function(x) {  
 # x is a list of values  
 min\_tmp <- min(x, na.rm = T)  
 max\_tmp <- max(x, na.rm = T)  
 mdn\_tmp <- median(x, na.rm = T)  
 mn\_tmp <- mean(x, na.rm = T)  
 rnge\_tmp <- range(x, na.rm = T)  
 sd\_tmp <- sd(x, na.rm = T)  
 cv\_tmp <- (sd\_tmp/mn\_tmp) \* 100  
 print(paste("Minimum =", min\_tmp, sep = " "))  
 print(paste("Maximum =", max\_tmp, sep = " "))  
 print(paste("Median =", mdn\_tmp, sep = " "))  
 print(paste("Mean =", mn\_tmp, sep = " "))  
 print(rnge\_tmp)  
 print(paste("Standard Deviation =", sd\_tmp, sep = " "))  
 print(paste("Coefficient of Variance =", cv\_tmp, sep = " "))  
   
}