Homework Assignment 1

Brendan Smith

January 28, 2016

Objective Statement: In this homework assignment, the goal is to familiarize ourselves with RStudio and RMarkdown given a dataset. We will apply some fundamental statistical analysis and plots utilizing built in R functions. The dataset provided is extracted from LiDAR data collected in the Sierra Nevada mountain range. Tree height and crown radii data was extracted through post processing. We are to import the data and test against the null hypothesis: Given tree height and crown radii, there will be zero correlation between the two.

Methods: First, the data is imported via the built-in `read.csv()` function. Next, a series data diagnostics are run to gather a general idea of the dataset, followed by the generation of histograms for the height and crown radii of the trees. Finally, the correlation between the tree height and crown radii is verified and presented.

Data: The data utilized in this analysis was collected using Light Detection and Ranging (LiDAR) of the conifers that reside in the Sierra Nevada mountain range. The tree height, crown-radii, and location were exctracted from the initial LiDAR dataset.

Code:

```
#Read in the values from Trees.csv and place them in the variable "Trees"
Trees <- read.csv("./Trees.csv",header = TRUE,sep = ",")
#Use the attach() function in order to access the data arrays imported from Trees.csv with their inheri
attach(Trees)
#Assemble a data frame with the variables for good house-keeping purposes
Trees.data<-data.frame(OBJECTID, x, y, z.m., r.m.)
#Remove the variable "Trees"
rm(Trees)
```

```
#Print a summary for the crown radii and tree height. This will provide the miniumum, first quartile, m
#Summary of Crown Radii
summary(Trees.data$r.m.)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.1847  1.5020  2.2420  2.5960  3.3190 12.0000
```

```
#Summary of Tree Height
summary(Trees.data$z.m.)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   2.000   6.155  10.900  13.910  17.780  73.930
```

In addition to analyzing the provided dataset, we are required to create a function in order to output key data diagnostics. Below is the function:

```
func = function(x) {
  #perform the required functions
  minim = min(x)
  avg = mean(x)
```

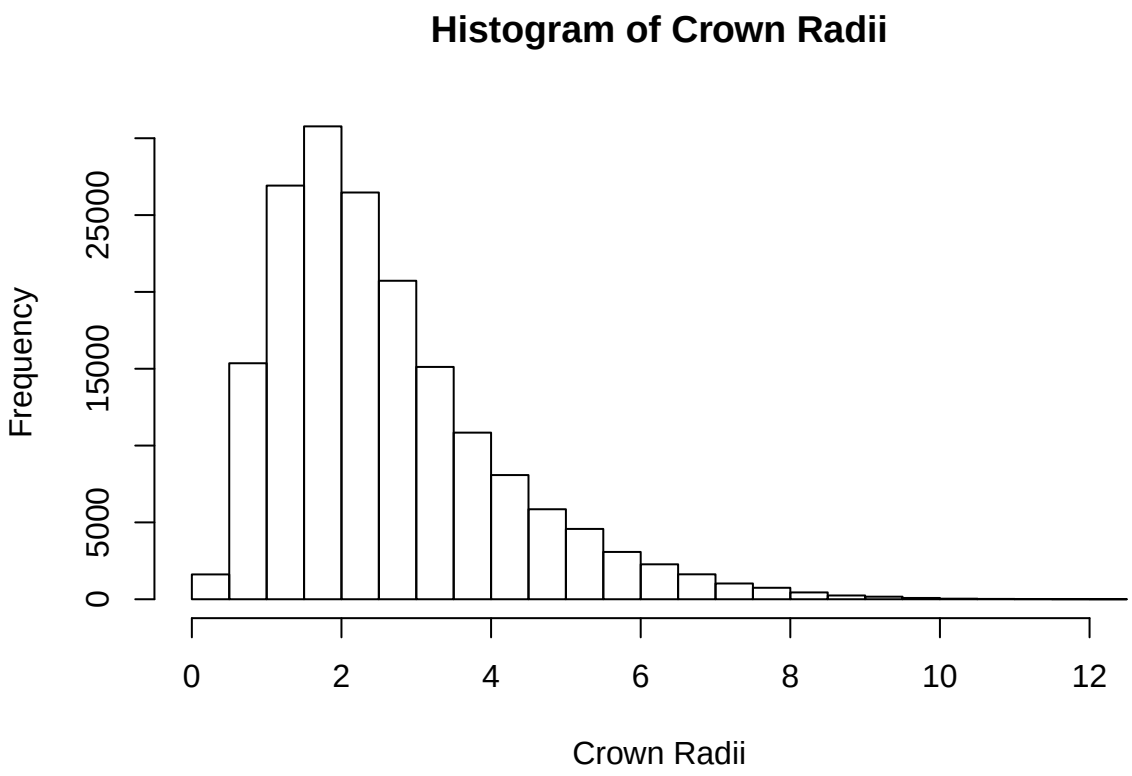
```
  med = median(x)
  maxi = max(x)
  ran = range(x)
  stan = sd(x)
  cv = stan/avg

  info <- c(minim, avg, med, maxi, ran, stan, cv)

  #print out all the calculated values in a human-readable way
  cat("Minimum: ", minim, "\n")
  cat("Mean: ", avg, "\n")
  cat("Median: ", med, "\n")
  cat("Maximum: ", maxi, "\n")
  cat("Range: ", ran, "\n")
  cat("Standard Deviation: ", stan, "\n")
  cat("Coefficient of Variation: ", cv, "\n")


  return(info) #return the minimum, average, median, maximum, range and standard deviation of the datas
}

#Test the functionality by inputing the tree height data
func(Trees.data$z.m.)
```

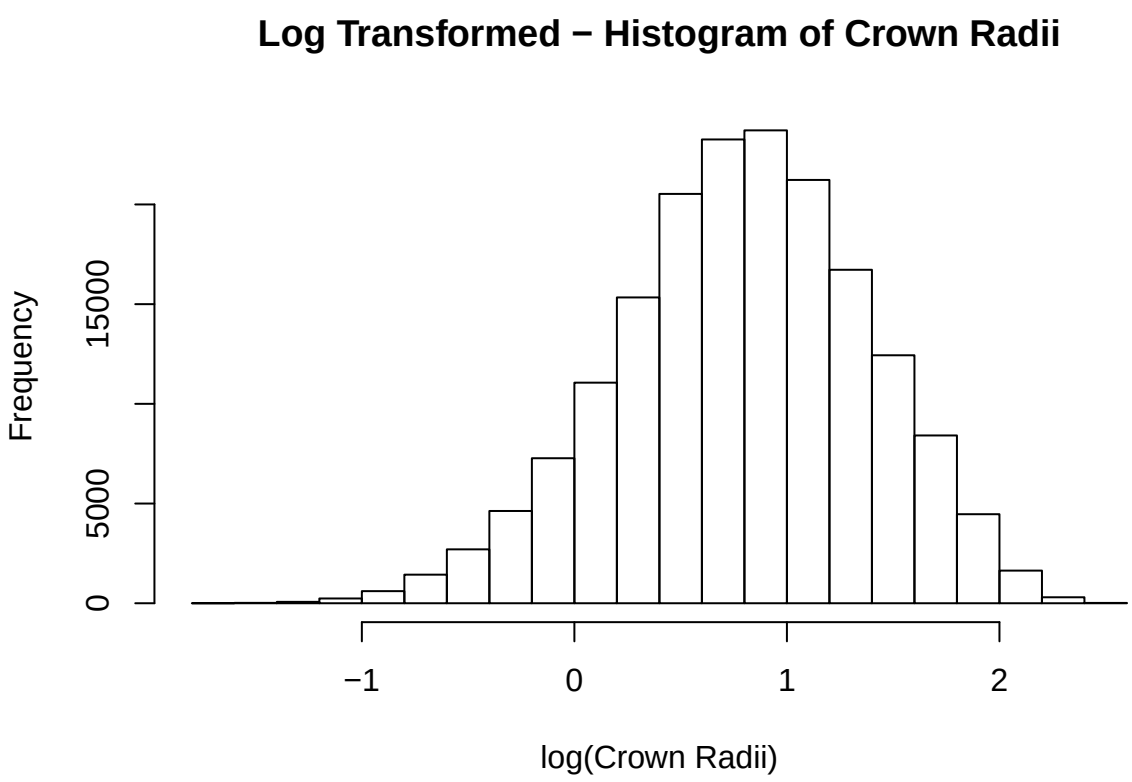
```
## Minimum:  2.0001
## Mean:  13.90983
## Median:  10.9011
## Maximum:  73.9267
## Range:  2.0001 73.9267
## Standard Deviation:  10.77207
## Coefficient of Variation:  0.7744215
```

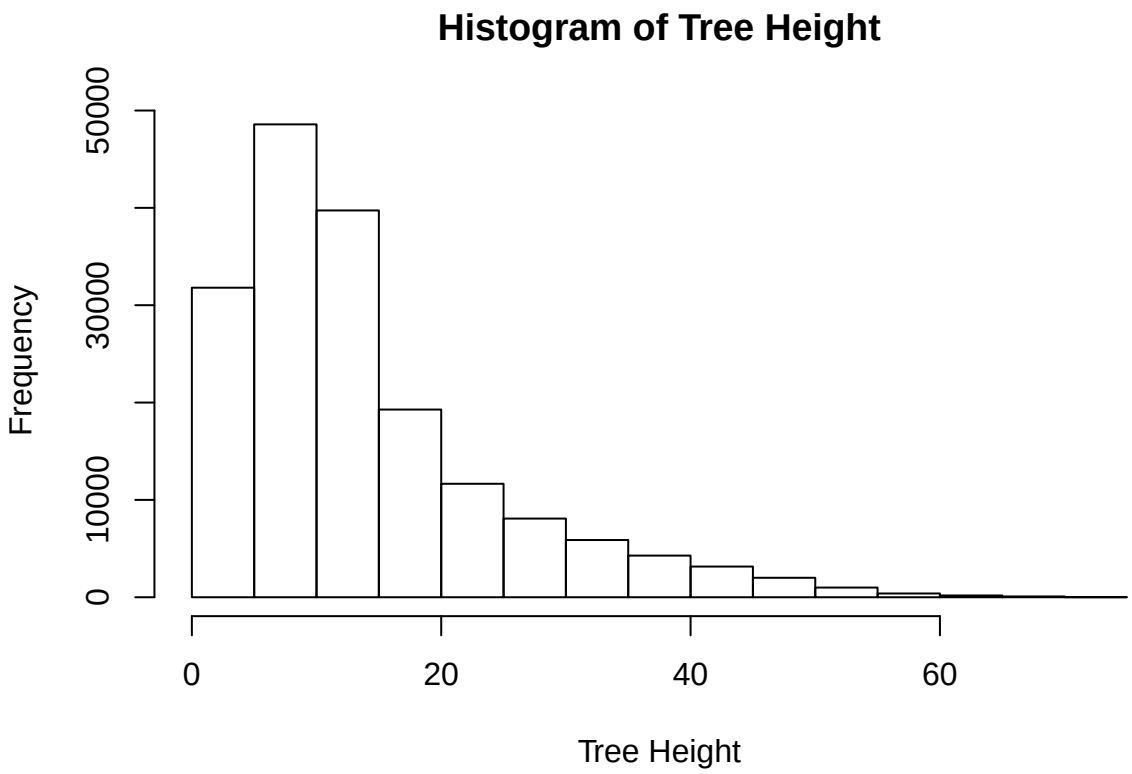
```
## [1]  2.0001000 13.9098253 10.9011000 73.9267000  2.0001000 73.9267000
## [7] 10.7720682  0.7744215
```

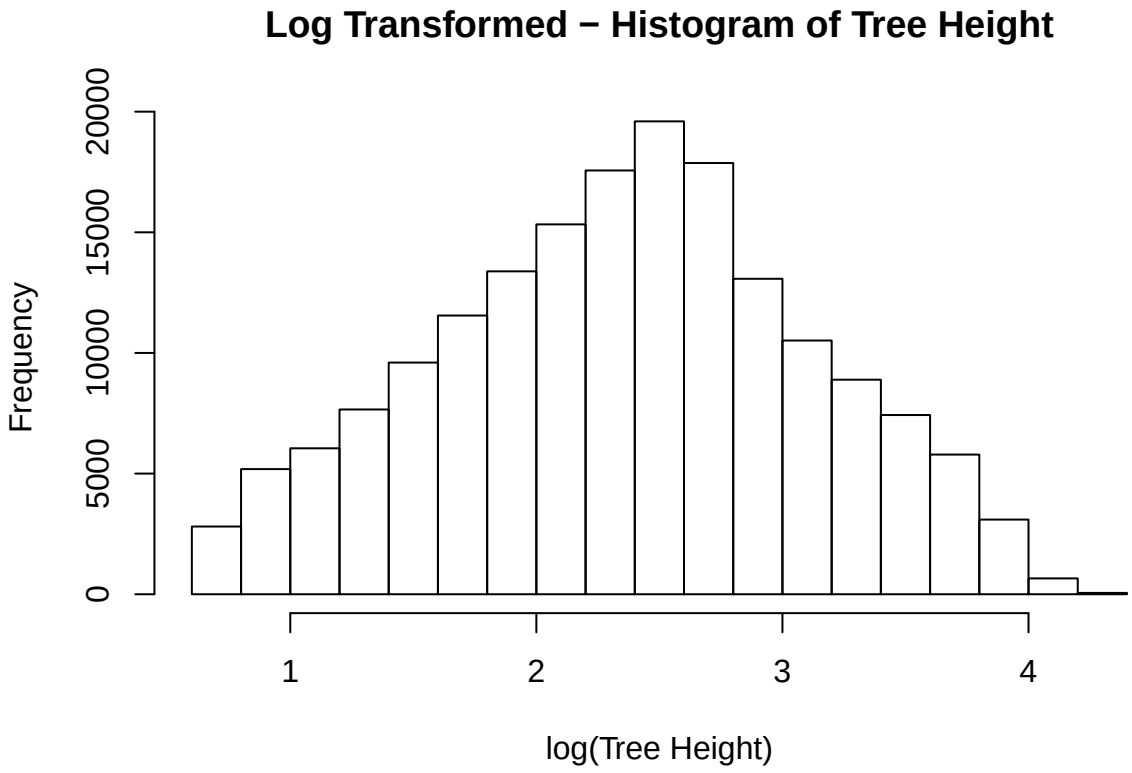
The function `func()` takes in a dataset, prints key data information in a human-readable form to the console, and also returns that data in an array.

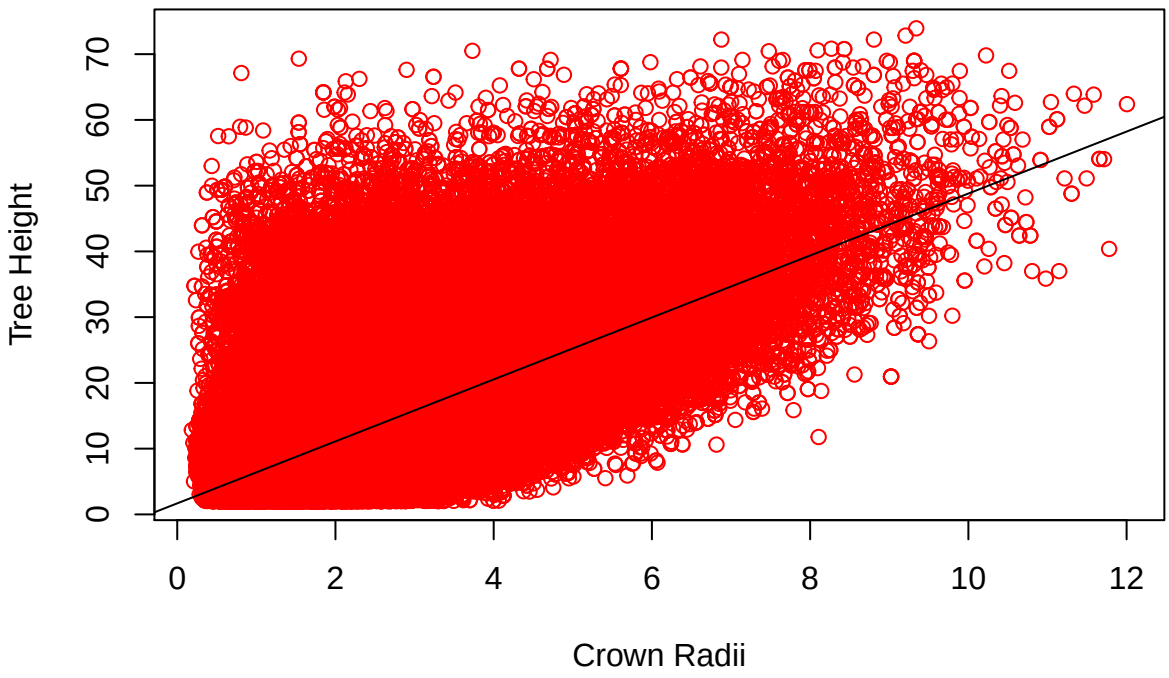
Results:

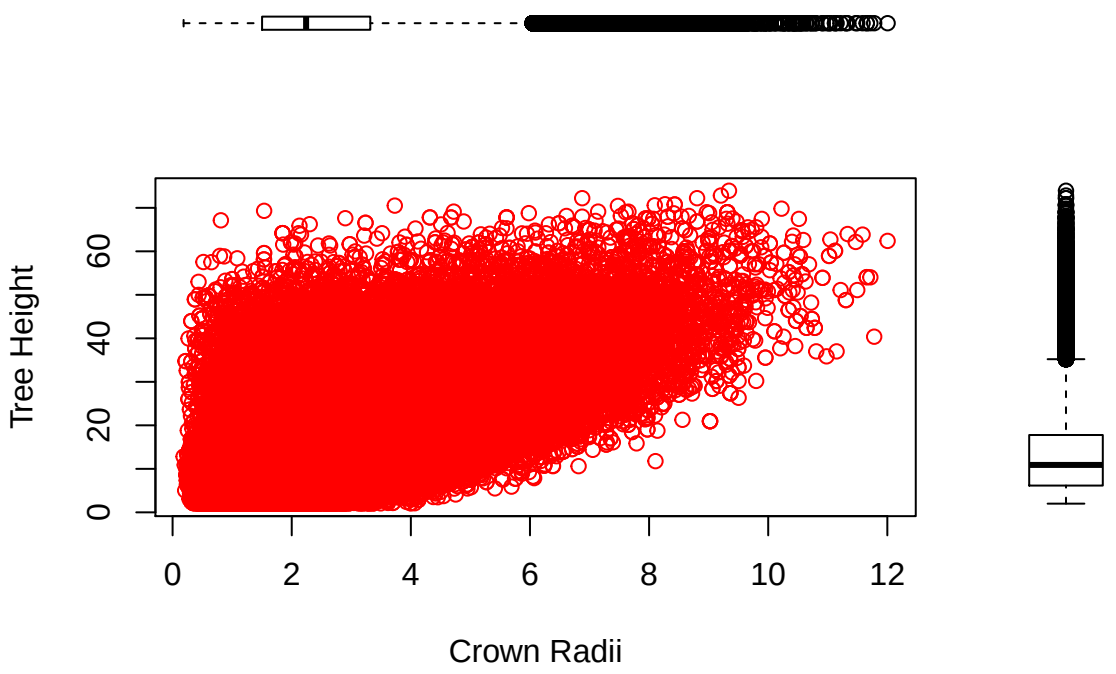

Histogram of Crown Radii
Frequency
0
5000
15000
25000
0
2
4
6
8
10
12
Crown Radii


Log Transformed − Histogram of Crown Radii
Frequency
0
5000
15000
−1
0
1
2
log(Crown Radii)


Histogram of Tree Height
Frequency
0
10000
30000
50000
0
20
40
60
Tree Height


Log Transformed − Histogram of Tree Height
Frequency
0
5000
10000
15000
20000
1
2
3
4
log(Tree Height)


Tree Height
0
10
20
30
40
50
60
70
0
2
4
6
8
10
12
Crown Radii



Discussion:

The uncorrected histograms indicate a left skewed distribution of tree height and tree-crown radii. Once the values are corrected using a log transformation, the histograms take on a normal distribution. This analysis shows just how powerful data transformations can be when analyzing data.

Plotting the tree height against the tree-crown radii with a scatterplot offers good qualitative insight to the correlation between tree height and tree-crown radii. Further analysis through the addition of a fitted line shows that the data lies along the `x=y` line. The addition of boxplots along the axes allows the viewer to visualize the distribution of each dataset while simultaneously visualizing the correlation.

The calculated Pearson's correlation coefficient, or r-value, is 0.6613456. This means that we have a strong positive correlation between the tree-crown radii and height, which is to be expected. This also confirms that we can reject the null hypothesis that there is no correlation between tree-crown radii and height.

Limitations: The most notable limitations to RMarkdown is the fact that you cannot underline or right-justify text. In terms of the data, the crown density is not taken into account when performing this analysis, and therefore could be considered an incomplete analysis.