

## assignment 4

Oyedayo Oyelowo

2 November 2017

```
#clear memory
rm(list = ls())
setwd("C:/Users/oyeda/Desktop/R_COURSE/assignment4")
#load the data
data<-read.table("puudata.txt", header = T, sep = "\t")
```

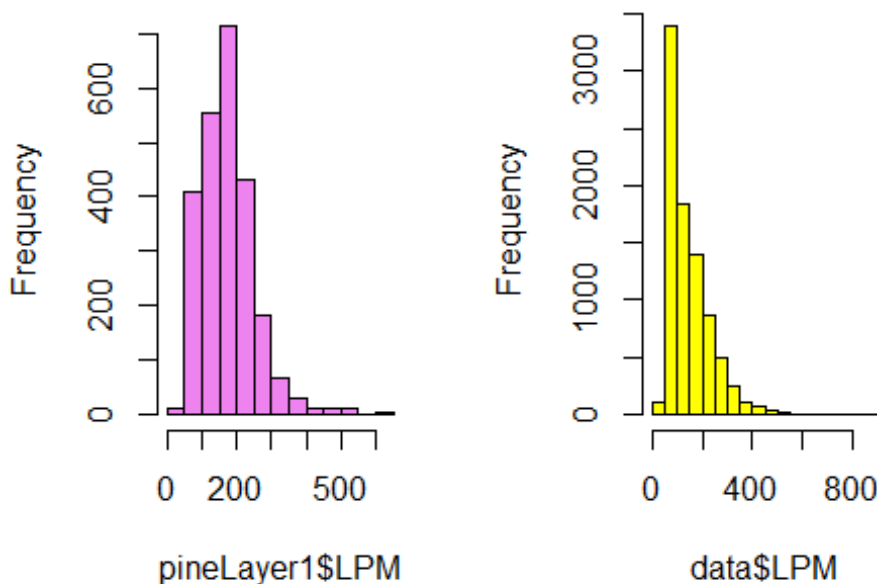
### Exercise 1

- Form separate histograms for diameters (LPM) of spruces and pines (tree species = PUULAJI, pine = 1 and spruce = 2) in the first crown layer (canopy cover layer = LATVKERROS).
- Which of the two is visually closer to normal distribution?

```
par(mfcol=c(1,2))
pineLayer1<- subset(data, data$PUULAJI==1 & data$LATVKERROS==1)
pine.hist<- hist(pineLayer1$LPM, main = "Histogram of diameter of Pine",
                 col = "violet")

spruceLayer1<-subset(data, data$PUULAJI==2 & data$LATVKERROS==1)
spruce.hist<-hist(data$LPM, main = "Histogram of diameter of Spruce",
                  col="yellow")
```

### Histogram of diameter of histogram of diameter of Spr

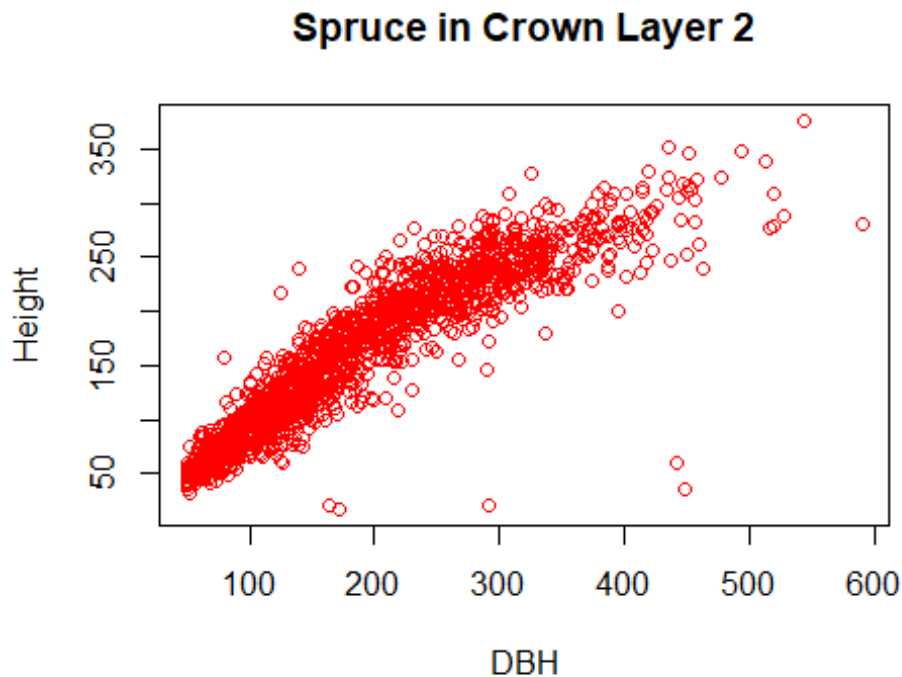


**Answer: Pine appears to be more normally distributed**

## Exercise 2

- Plot tree height (PITUUS) as function of tree diameter (LPM) (height on the y-axis and dbh on the x-axis)
  - for spruces in the first canopy cover layer

```
plot(spruceLayer1$LPM, spruceLayer1$PITUUS,  
     xlab = "DBH", ylab = "Height", main="Spruce in Crown Layer 2", col="red")
```

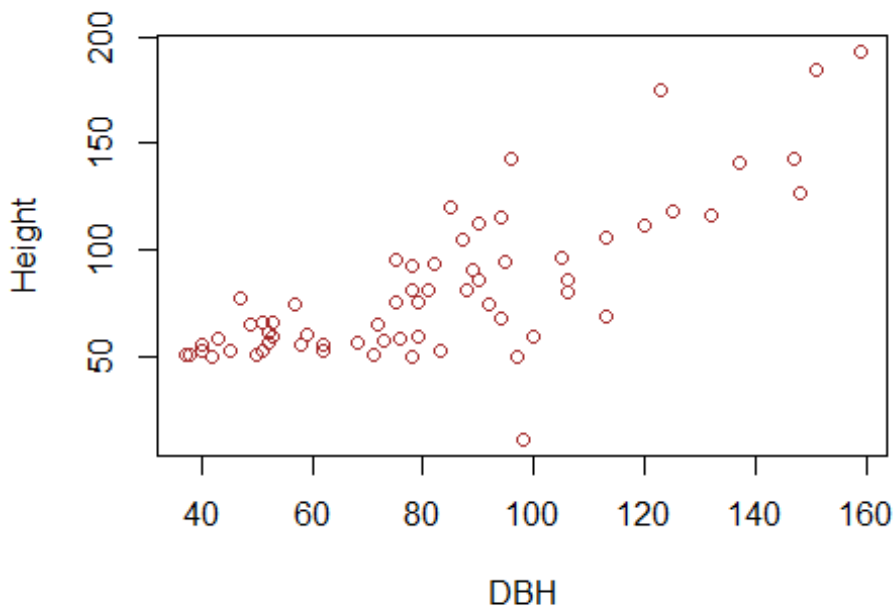


```
#plot(x=spruceLayer1$LPM, y=spruceLayer1$PITUUS) alternatively
```

- for pines in the second canopy cover layer

```
pineLayer2<- subset(data, data$PUULAJI==1 & data$LATVKERROS==2)  
plot(pineLayer2$PITUUS, pineLayer2$LPM, main = "Pines in Crown Layer 2",  
     xlab = "DBH", ylab = "Height", col="brown")
```

## Pines in Crown Layer 2

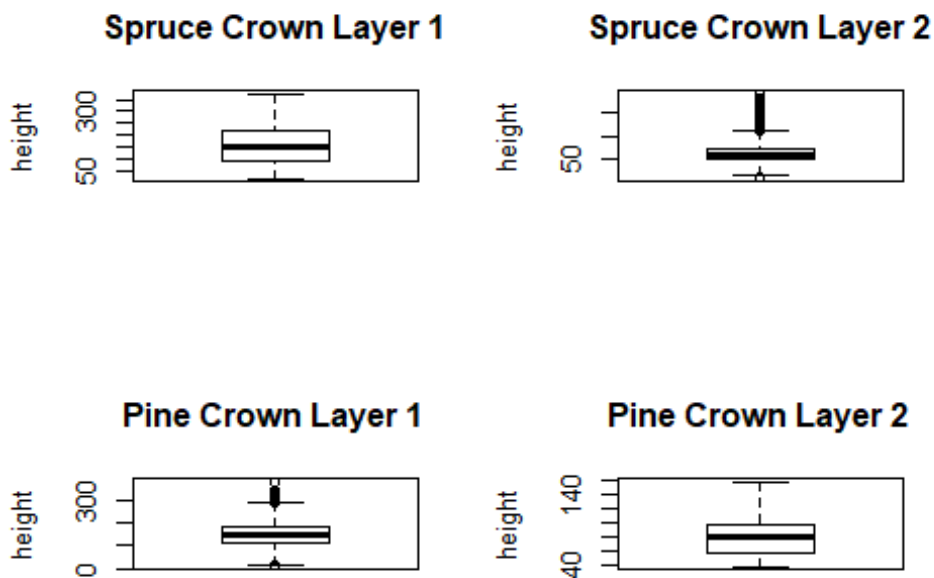


### Exercise 3

- Form a box-and-whiskers plot for heights of conifers (species 1 and 2) in first and second crown layer.
- What can you tell from the difference between height distributions of pines and spruces?

```
par(mfrow=c(2,2), bg="white")
boxplot(spruceLayer1$PITUUS, ylab="height", main="Spruce Crown Layer 1")
spruceLayer2<-subset(data, data$PUULAJI==2 & data$LATVKERROS==2)
boxplot(spruceLayer2$PITUUS, ylab="height", main="Spruce Crown Layer 2")

boxplot(pineLayer1$PITUUS, ylab="height", main="Pine Crown Layer 1")
boxplot(pineLayer2$PITUUS, ylab="height", main="Pine Crown Layer 2")
```

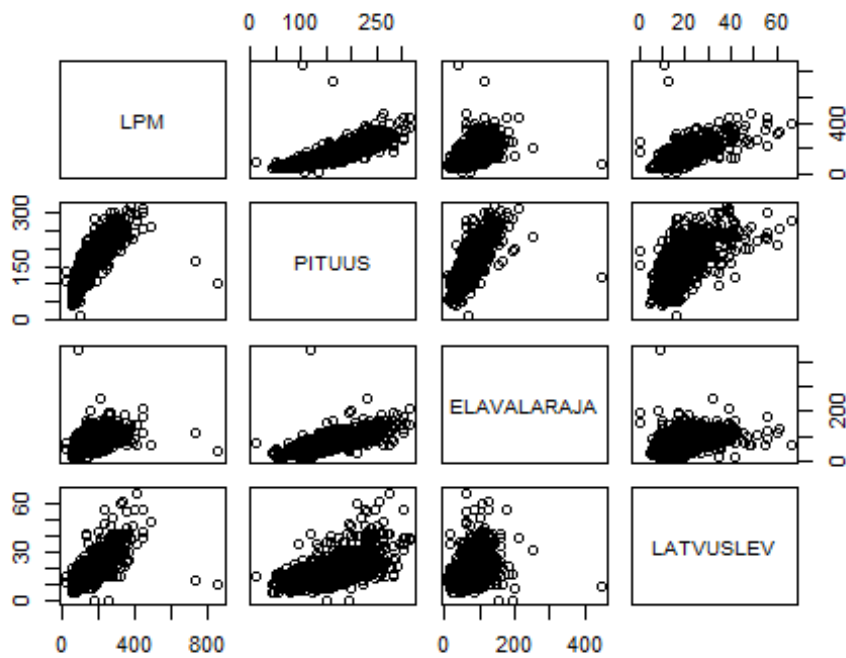


*The Spruce in the crown layer one appears to be more varying and less varying in the second crown layer. Pine, on the other hand, is less varying in the first crown layer but more varying in the second crown layer.*

#### Exercise 4

- Plot the diameter, height, height of living crown (ELAVALARAJA), and width of the crown (LATVUSLEV) for silver birches (PUULAJI = 3) in the first canopy cover layer with pairs() function.
- Choose the two variables that seem to have the most linear relationship.
- Examine the distribution of these two variables by drawing a QQ-plot.
- What can you tell from the plot?

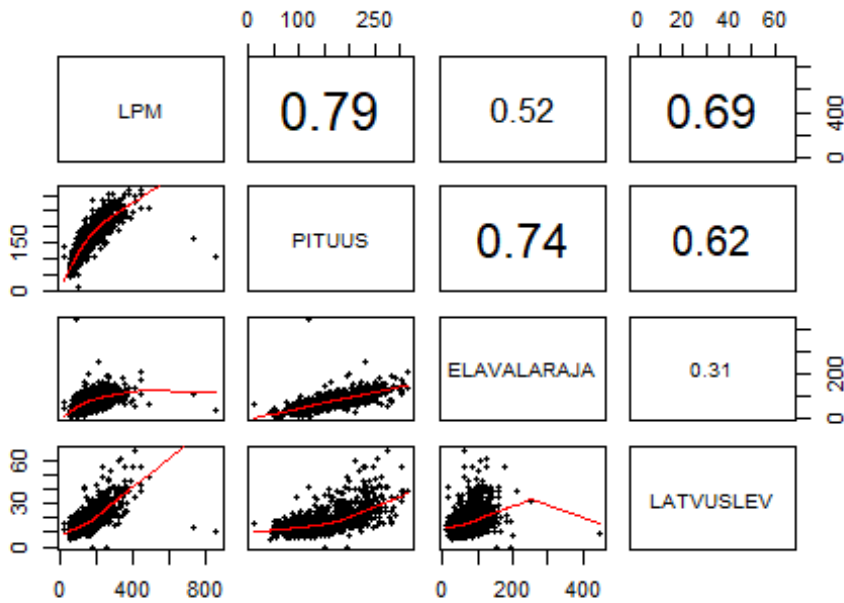
```
s.birchLayer1<- subset(data, data$PUULAJI==3 & data$LATVKERROS==1)
#simple pairing but I used a more detailed pairer after this
pairs(~LPM + PITUUS + ELAVALARAJA + LATVUSLEV, data = s.birchLayer1)
```



```
## put (absolute) correlations on the upper panels,
## with size proportional to the correlations.
panel.cor <- function(x, y, digits = 2, prefix = "", cex.cor, ...)
{
  usr <- par("usr"); on.exit(par(usr))
  par(usr = c(0, 1, 0, 1))
  r <- abs(cor(x, y))
  txt <- format(c(r, 0.123456789), digits = digits)[1]
  txt <- paste0(prefix, txt)
  if(missing(cex.cor)) cex.cor <- 0.8/strwidth(txt)
  text(0.5, 0.5, txt, cex = cex.cor * r)
}

pairs(~LPM+PITUUS+ELAVALARAJA+LATVUSLEV, data=s.birchLayer1,
      lower.panel=panel.smooth, upper.panel=panel.cor,
      pch=20, main="Birch Scatterplot Matrix")
```

## Birch Scatterplot Matrix



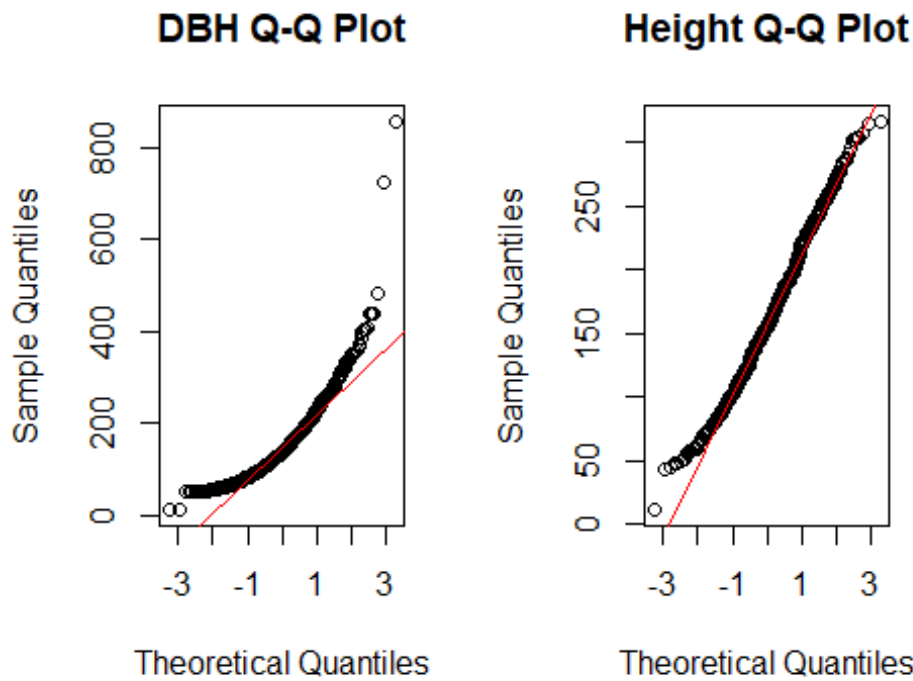
**diameter(LPM) and height(PITUUS), have the most linear relationship with a correlation of 0.79**

- Plot 3: similar plot using ggplot2

*#?qqplot*

```
par(mfrow=c(1,2))
qqnorm(s.birchLayer1$LPM, main = "DBH Q-Q Plot")
qqline(s.birchLayer1$LPM, col="red")

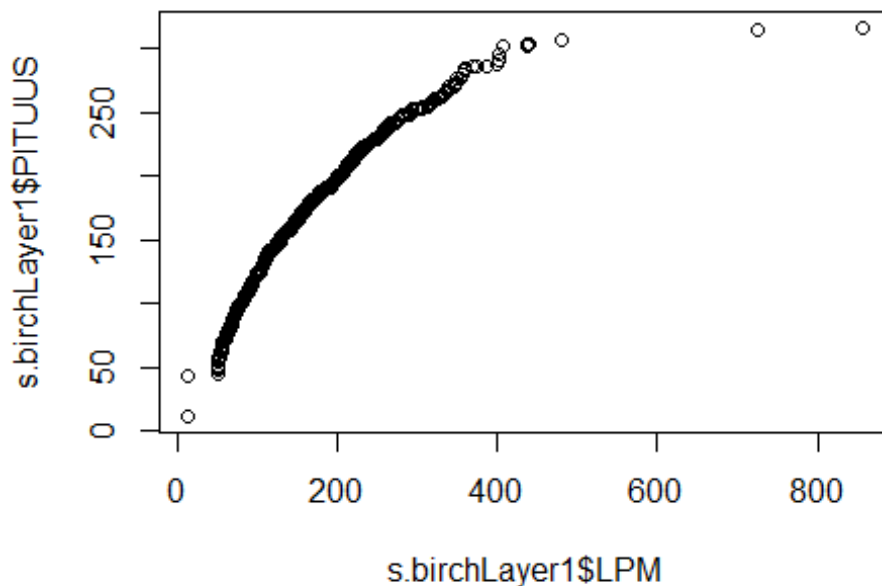
qqnorm(s.birchLayer1$PITUUS, main = "Height Q-Q Plot")
qqline(s.birchLayer1$PITUUS, col="red")
```



diameter(LPM) is not normally distributed but skewed to the right, height appears to be normally distributed

- Although, The Shapiro-Wilk test for normality is used to examine a continuous variable, I decided to apply it here too.

```
a<-qqplot(s.birchLayer1$LPM, s.birchLayer1$PITUUS, plot.it = TRUE,
  xlab = deparse(substitute(s.birchLayer1$LPM)),
  ylab = deparse(substitute(s.birchLayer1$PITUUS)))
```



```
##?shapiro.test
shapiro.test(s.birchLayer1$LPM)

##
##  Shapiro-Wilk normality test
##
## data:  s.birchLayer1$LPM
## W = 0.87534, p-value < 2.2e-16

shapiro.test(s.birchLayer1$PITUUS)

##
##  Shapiro-Wilk normality test
##
## data:  s.birchLayer1$PITUUS
## W = 0.99075, p-value = 1.52e-05
```

**based on the fact that the p values of the height and DBH(LPM) are less than 0.05, we can reject the null hypothesis that the data the data are normally distributed**