Selina Narain DTSC 630 Homework 2

2023-02-24

## Question 1

**The dataframe rainforest from DAAG package has data on several rainforest species.**

**Explore the dataframe rainforest using str, summary, table functions.**

library(DAAG)  
data("rainforest")  
str(rainforest)

## 'data.frame': 65 obs. of 7 variables:  
## $ dbh : num 6 23 20 23 24 5 5 8 10 8 ...  
## $ wood : num NA 353 208 445 590 14 10 31 59 30 ...  
## $ bark : num NA NA NA NA NA NA NA NA NA NA ...  
## $ root : num 6 135 NA NA NA 2 NA NA NA 6 ...  
## $ rootsk : num 0.3 13 NA NA NA 2.4 NA NA NA 1 ...  
## $ branch : num NA 35 41 50 NA NA NA NA NA 4 ...  
## $ species: Factor w/ 4 levels "Acacia mabellae",..: 1 1 1 1 1 1 1 1 1 1 ...

summary(rainforest)

## dbh wood bark root   
## Min. : 4.00 Min. : 3.0 Min. : 8.00 Min. : 2.00   
## 1st Qu.: 8.00 1st Qu.: 29.0 1st Qu.: 11.75 1st Qu.: 6.00   
## Median :12.00 Median : 100.0 Median : 45.50 Median : 16.00   
## Mean :16.06 Mean : 265.4 Mean : 51.00 Mean : 30.85   
## 3rd Qu.:22.00 3rd Qu.: 386.5 3rd Qu.: 84.75 3rd Qu.: 44.00   
## Max. :56.00 Max. :1530.0 Max. :105.00 Max. :135.00   
## NA's :1 NA's :61 NA's :52   
## rootsk branch species   
## Min. : 0.300 Min. : 4.00 Acacia mabellae:16   
## 1st Qu.: 1.300 1st Qu.: 9.00 C. fraseri :12   
## Median : 2.400 Median : 25.00 Acmena smithii :26   
## Mean : 7.477 Mean : 32.86 B. myrtifolia :11   
## 3rd Qu.:13.000 3rd Qu.: 45.50   
## Max. :24.000 Max. :120.00   
## NA's :52 NA's :22

#table(rainforest)

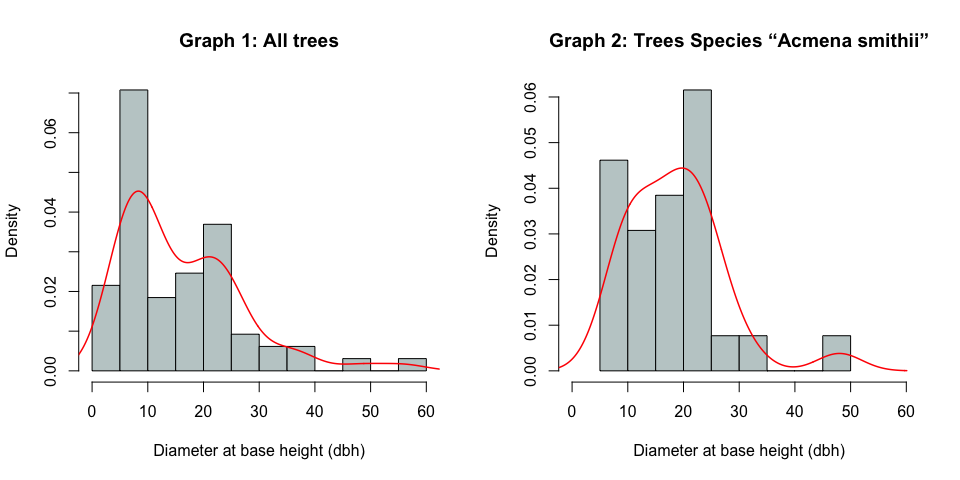
**From the above exploration, what are the names and numbers of the species in the data set?**

The names and numbers of the species in the data set are: Acacia mabellae: 16 C. fraseri: 12  
Acmena smithii: 26 B. myrtifolia: 11

**Plot two side-by-side graphs:**   
the left of which is  
– a histogram showing the distribution of the variable dbh (diameter at base height) for all the trees;  
– superimposed with a density plot (which estimates the number of values per unit interval) using density() function.

The graph on the right is a similar histogram with fitted density curve for trees of species “Acmena smithii”.

#Setting the plot area to a 1 by 2 array. One row of plots, and two columns.  
par(mfrow = c(1,2))  
  
#Left graph  
hist(rainforest$dbh,  
 main="Graph 1: All trees",  
 freq = FALSE,  
 xlab = "Diameter at base height (dbh)",  
 col = "azure3",  
 prob = TRUE,  
 border = "black")  
#Density plot  
lines(density(rainforest$dbh), col = 'red', lwd = 1.5)  
  
#Takes subset of Acmena smithii species data and load into dataframe  
df\_acmena <- subset(rainforest, species=="Acmena smithii")  
  
#Right graph  
hist(df\_acmena$dbh,  
 main="Graph 2: Trees Species “Acmena smithii”",  
 xlab = "Diameter at base height (dbh)",  
 xlim = c(0, 60),  
 col = "azure3",  
 prob = TRUE,  
 border = "black")   
  
#Density plot  
lines(density(df\_acmena$dbh), col = 'red', lwd = 1.5)



## Question 2

**Use the mtcars dataset, provided by base R to plot 3 scatter plots in the same graph.** **The three pair of data are (mpg, drat), (mpg, wt), (mpg, qsec).**  **You can start with the main plot using plot function (mpg, drat, …), then add/superimpose the other two plots by points() as follows:**  
***plot(mpg, drat, …)***  
***points(mpg, wt, …)***  
***points(mpg, qsec, …)***

data(mtcars)  
head(mtcars)

## mpg cyl disp hp drat wt qsec vs am gear carb  
## Mazda RX4 21.0 6 160 110 3.90 2.620 16.46 0 1 4 4  
## Mazda RX4 Wag 21.0 6 160 110 3.90 2.875 17.02 0 1 4 4  
## Datsun 710 22.8 4 108 93 3.85 2.320 18.61 1 1 4 1  
## Hornet 4 Drive 21.4 6 258 110 3.08 3.215 19.44 1 0 3 1  
## Hornet Sportabout 18.7 8 360 175 3.15 3.440 17.02 0 0 3 2  
## Valiant 18.1 6 225 105 2.76 3.460 20.22 1 0 3 1

str(mtcars)

## 'data.frame': 32 obs. of 11 variables:  
## $ mpg : num 21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...  
## $ cyl : num 6 6 4 6 8 6 8 4 4 6 ...  
## $ disp: num 160 160 108 258 360 ...  
## $ hp : num 110 110 93 110 175 105 245 62 95 123 ...  
## $ drat: num 3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ...  
## $ wt : num 2.62 2.88 2.32 3.21 3.44 ...  
## $ qsec: num 16.5 17 18.6 19.4 17 ...  
## $ vs : num 0 0 1 1 0 1 0 1 1 1 ...  
## $ am : num 1 1 1 0 0 0 0 0 0 0 ...  
## $ gear: num 4 4 4 3 3 3 3 4 4 4 ...  
## $ carb: num 4 4 1 1 2 1 4 2 2 4 ...

summary(mtcars)

## mpg cyl disp hp   
## Min. :10.40 Min. :4.000 Min. : 71.1 Min. : 52.0   
## 1st Qu.:15.43 1st Qu.:4.000 1st Qu.:120.8 1st Qu.: 96.5   
## Median :19.20 Median :6.000 Median :196.3 Median :123.0   
## Mean :20.09 Mean :6.188 Mean :230.7 Mean :146.7   
## 3rd Qu.:22.80 3rd Qu.:8.000 3rd Qu.:326.0 3rd Qu.:180.0   
## Max. :33.90 Max. :8.000 Max. :472.0 Max. :335.0   
## drat wt qsec vs   
## Min. :2.760 Min. :1.513 Min. :14.50 Min. :0.0000   
## 1st Qu.:3.080 1st Qu.:2.581 1st Qu.:16.89 1st Qu.:0.0000   
## Median :3.695 Median :3.325 Median :17.71 Median :0.0000   
## Mean :3.597 Mean :3.217 Mean :17.85 Mean :0.4375   
## 3rd Qu.:3.920 3rd Qu.:3.610 3rd Qu.:18.90 3rd Qu.:1.0000   
## Max. :4.930 Max. :5.424 Max. :22.90 Max. :1.0000   
## am gear carb   
## Min. :0.0000 Min. :3.000 Min. :1.000   
## 1st Qu.:0.0000 1st Qu.:3.000 1st Qu.:2.000   
## Median :0.0000 Median :4.000 Median :2.000   
## Mean :0.4062 Mean :3.688 Mean :2.812   
## 3rd Qu.:1.0000 3rd Qu.:4.000 3rd Qu.:4.000   
## Max. :1.0000 Max. :5.000 Max. :8.000

#plot(mtcars$drat)  
#plot(mtcars$wt)  
#plot(mtcars$qsec)  
  
#Main plot using plot function (mpg, drat)  
plot(mtcars$mpg,  
 mtcars$drat,   
 main = "Scatter Plot (mpg, drat), (mpg, wt), (mpg, qsec)",   
 xlab = "Miles per gallon (mpg)",   
 ylab = "Rear axle ratio (drat)",  
 xlim = c(0, 35),  
 ylim = c(0, 25),  
 pch = 16,  
 col = "deeppink3")  
#Add/superimpose the other two plots by points() for (mpg, wt), (mpg, qsec)  
points(mtcars$mpg,  
 mtcars$wt,   
 pch = 16,  
 col = "yellowgreen")  
points(mtcars$mpg,  
 mtcars$qsec,   
 pch = 16,  
 col = "deepskyblue")  
  
#Creating legend (colors, symbols, labels)  
leg\_cols <- c("deeppink3", "yellowgreen", "deepskyblue")  
leg\_sym <- c(16, 16, 16)  
leg\_lab <- c("(mpg, drat)", "(mpg, wt)", "(mpg, qsec)")  
  
legend(x = 0, y = 25, col = leg\_cols, pch = leg\_sym, legend = leg\_lab, title = "Legend")

