For all of these solutions, first the data needs to be loaded

>> tree_data = read.table("TreeData.txt", header=TRUE)

1) Produce a frequency distribution for the variable "Type" and "Nest"

>> table(tree_data$Type)

>> table(tree_data$Nest)

Type:	A	B	C	C
	46	40	44	30

Nest:	1	2	3	4	5
	41	41	27	26	25

2) Produce a cumulative relative frequency distribution for "Nest". What percent of trees had less than two birds nests in them? What percent of trees had one or more birds nests but less than 4?
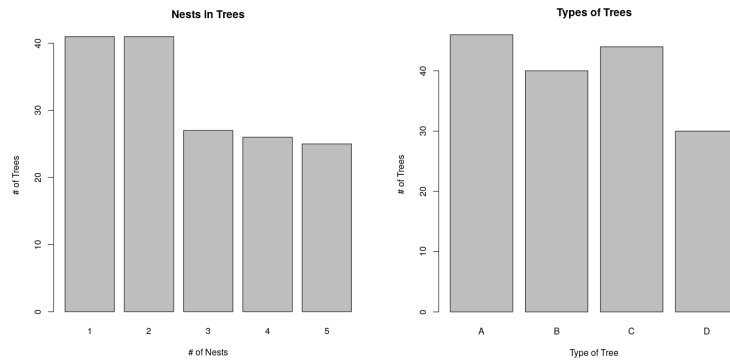
>> cumsum(table(tree_data$Nest)/length(tree_data$Nest))

Nest:	1	2	3	4	5
	0.25625	0.5125	0.68125	0.84375	1

From the table above we see that 25.625% of tree had less than two birds nests in them, and that 68.125% of trees had one or more birds nests but less than 4.

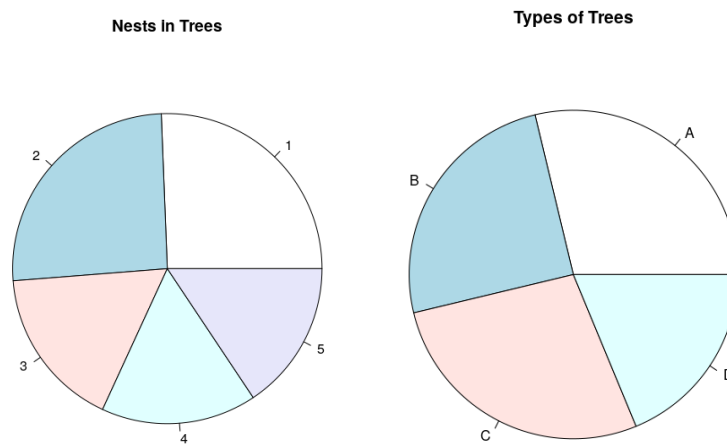**3)** Produce a bar chart and pie chart for "Nest" and "Type"

>> barplot(table(tree_data$Nest), main="Nests in Trees", xlab="# of Nests", ylab="# of Trees")

>> barplot(table(tree_data$Type), main="Types of Trees", xlab="Type of Tree", ylab=" of Trees")

>> pie(table(tree_data$Nest), labels=c(1,2,3,4,5), main="Nests in Trees")

>> pie(table(tree_data$Type), labels=c("A","B","C","D"), main="Types of Trees")

4) Produce the following descriptive statistics for "Height": mean, standard deviation, median, maximum value, minimum value, first quartile and third quartile.

>> mean(tree_data$Height)   >> cat("Minimum Value: ", quartiles[1])

>> sd(tree_data$Height)   >> cat("Maximum Value: ", quartiles[5])

>> tree_data$Height   >> cat("First Quartile: ", quartiles[2])

>> quantile(tree_data$Height)   >> cat("Third Quartile: ", quartiles[4])

>> median(tree_data$Height)

Mean: 39.08694   Std Dev: 6.07066

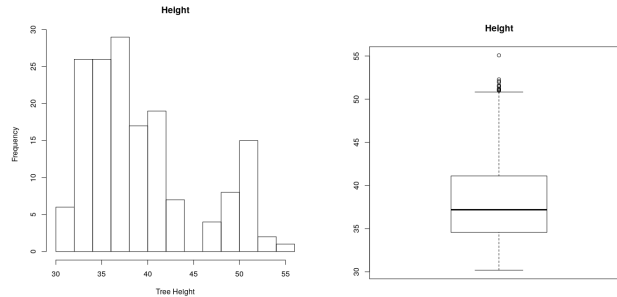Minimum Value: 30.18   Maximum Value: 55.08

First Quartile: 34.56   Third Quartile: 40.965

Median: 37.185

5) Produce a histogram and box plot of the variable "Height". Describe some of the characteristics of this data set using the histogram.

>> hist(tree_data$Height, main="Height", xlab="Tree Height")

>> boxplot(tree_data$Height, main="Height")

From the histogram of height, we can see that it is a bi-modal distribution, with most trees falling in the 32ft to 37ft range.
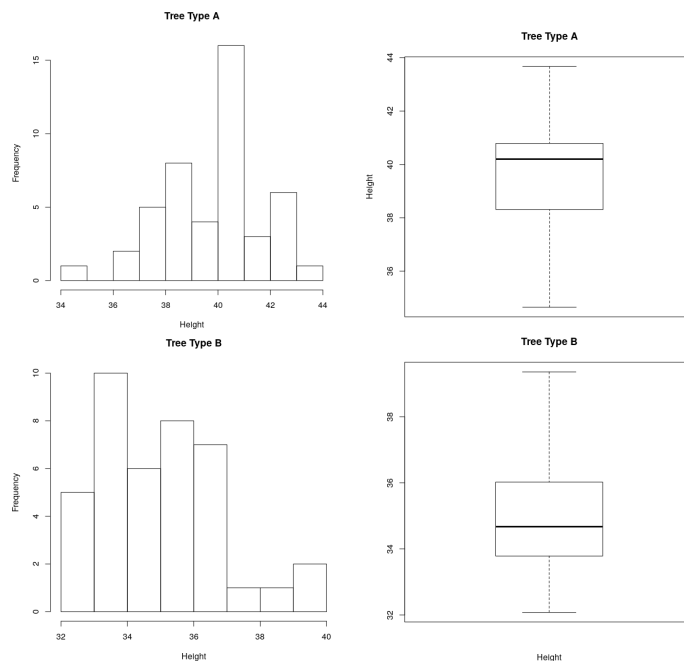
6) Produce a histogram and box plot of the variable "Height" for each of the tree types separately.
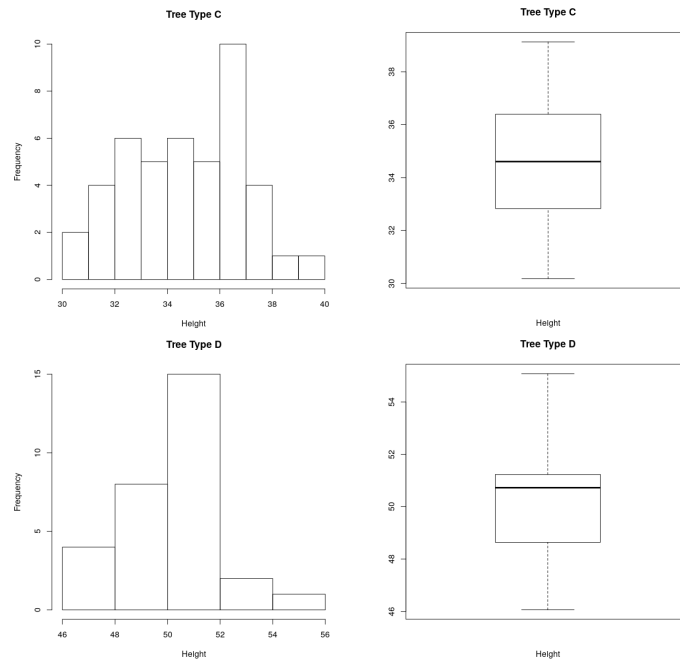
Replace "A" with "B", "C" and "D" to produce other graphs.

>> sub = subset(tree_data, tree_data$Type == "A")$Height

>> hist(sub, main="Tree Type A", xlab="Height")

>> boxplot(sub, main="Tree Type A", ylab="Height")

7) Produce a kernel density plot for "Height" for each of the different tree types. Overlay the plot of the kernel density with a plot of the theoretical normal distribution constructed using the estimated mean and standard deviation. Include a legend in the plot. Does the plot provide evidence the data is normally distributed?

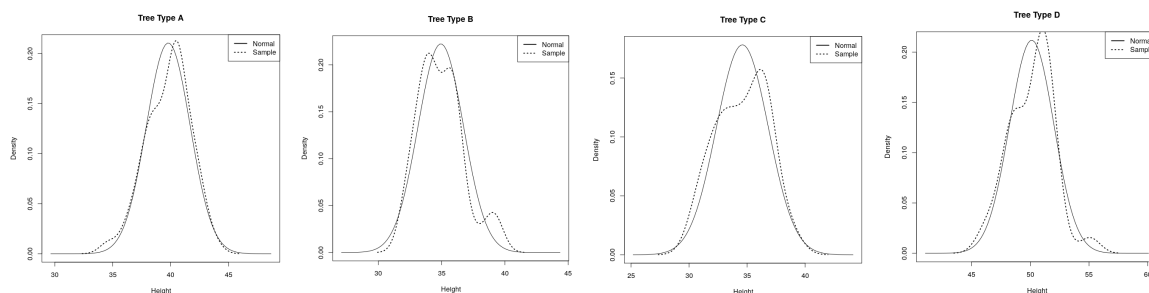Replace "A" with "B", "C" and "D" to produce other graphs.

\>> sub = subset(tree_data, tree_data\$Type == "A")\$Height

\>> curve(dnorm(x, mean(sub), sd(sub)), from=min(sub)-5,to=max(sub)+5, >> main="Tree Type A", xlab="Height", ylab="Density")

\>> lines(density(sub), type="l", lwd=2, lty=3)

\>> legend("topright", c("Normal", "Sample"), lty=c(1,3), lwd=2)

All of the plots provide evidence that the heights of each tree by type are normally distributed.

8) Implement the Shapiro-Wilks test for the variable "Height". Explain what conclusion the test is supporting and how you know. Does that conclusion support your conclusion in part 7?

  Replace "A" with "B", "C" and "D" to produce other graphs.

  >> sub = subset(tree_data, tree_data$Type == "A")$Height

  >> shapiro.test(sub)

Shapiro-Wilks Test Results:

Tree Type	Shapiro-Wilks P Value
A	0.5751
B	0.0423
C	0.3352
D	0.3346
All	9.215e-10

The results from the Shapiro Wilks Tests support that the distribution of each tree type by height are all fairly normal with type B being on the line of normality. Formally, with a significant p-value of .05, type B would fail the Shapiro-Wilks test for normality. The test results also show that is is highly unlikely that the entire set of tree heights is normally distributed.