BT Exploratory Analysis

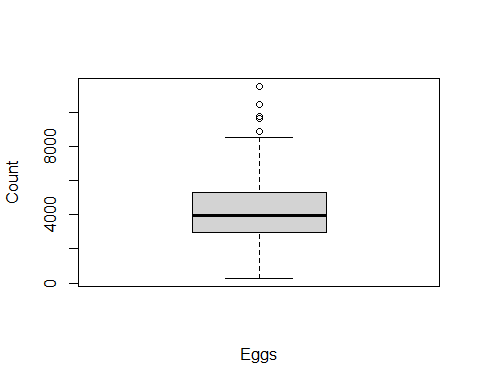
Craig Wells

March 2, 2021

#Load the dataset.  
BT=read.csv("BT.csv",header=TRUE)  
#Recall the column names.  
names(BT)

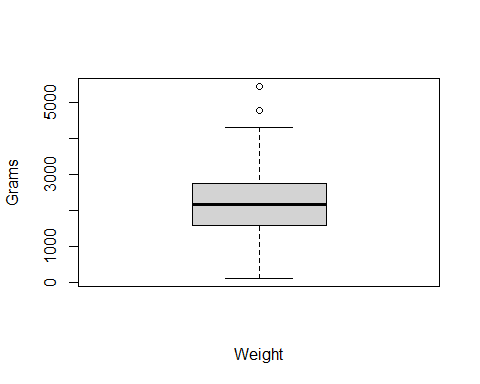
## [1] "Author" "Age" "TL" "FL" "WT" "Eggs"

#Generate a boxplot of the Eggs column.  
boxplot(BT$Eggs, xlab="Eggs", ylab="Count")



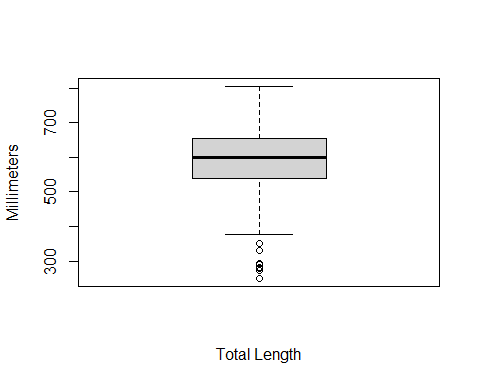
There are some outliers, but they should be true values.

#Generate a boxplot of the Weight column.  
boxplot(BT$WT, xlab="Weight", ylab="Grams")



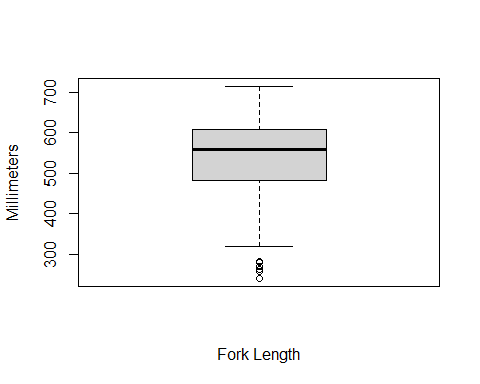
There are some outliers, but they should be true values.

#Generate a boxplot of the Total Length column  
boxplot(BT$TL, xlab="Total Length", ylab="Millimeters")



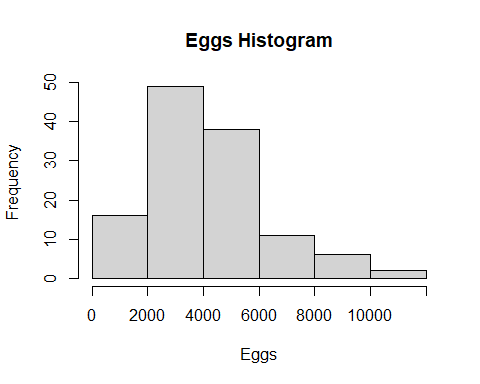
There are some outliers, but they should be true values.

#Generate a boxplot of the Fork Length column  
boxplot(BT$FL, xlab="Fork Length", ylab="Millimeters")

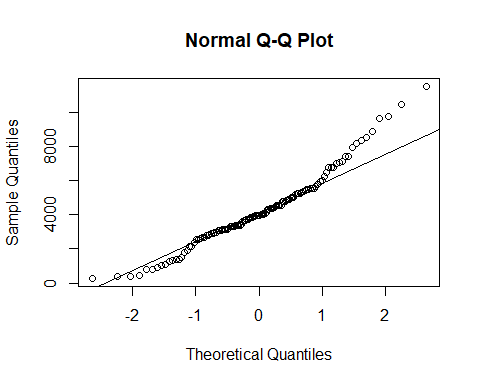


There are some outliers, but they should be true values.

#Generate a historgram of Eggs  
BT$Eggs=as.numeric(BT$Eggs)  
hist(BT$Eggs,xlab="Eggs",main="Eggs Histogram")

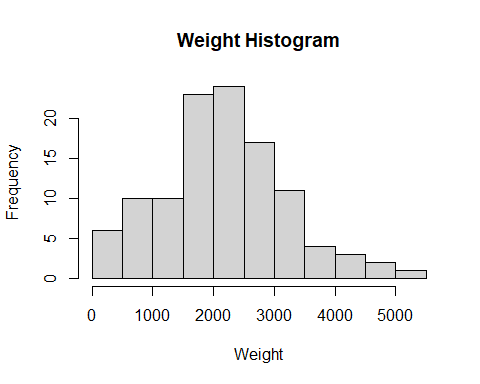


#Generate a QQ Plot and line for eggs.  
qqnorm(BT$Eggs)  
qqline(BT$Eggs)



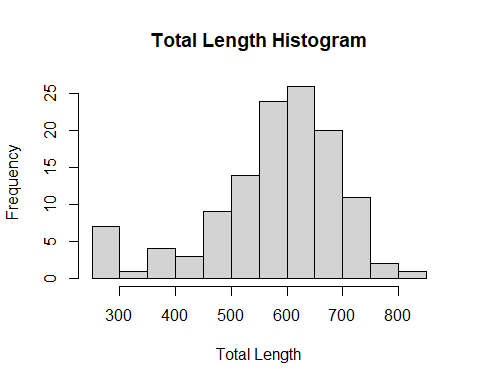
These data are not normally distributed.

#Generate a historgram of Weight  
hist(BT$WT,xlab="Weight",main="Weight Histogram")



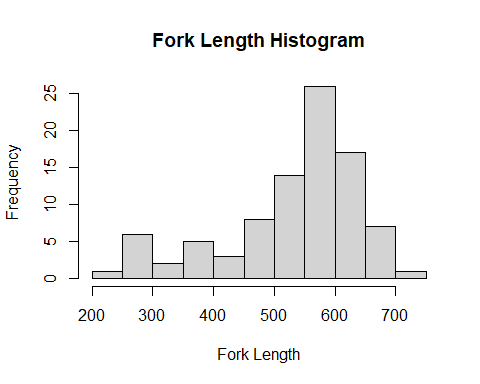
These data are not normally distributed.

#Generate a historgram of Total Length  
hist(BT$TL,xlab="Total Length",main="Total Length Histogram")



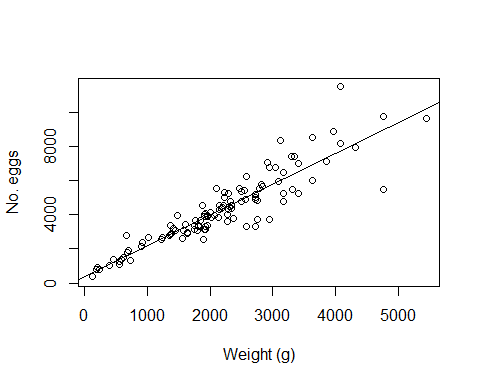
These data are not normally distributed.

#Generate a historgram of Fork Length  
hist(BT$FL,xlab="Fork Length",main="Fork Length Histogram")



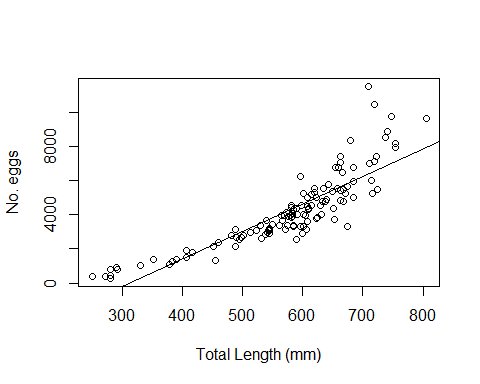
These data are not normally distributed

#Scatterplot of Eggs vs Weight  
plot(BT$WT,BT$Eggs,xlab="Weight (g)",ylab="No. eggs")  
  
#Linear regression for Eggs vs WT  
mymodel=lm(Eggs~WT,data=BT)  
abline(mymodel)



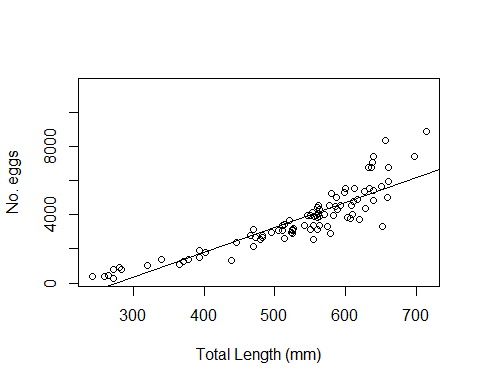
The relationship is fairly linear. Linear Regression seems to fit the points fairly well.

#Scatterplot of Eggs vs Total Length  
plot(BT$TL,BT$Eggs,xlab="Total Length (mm)",ylab="No. eggs")  
#Linear regression for Eggs vs TL  
mymodel2=lm(Eggs~TL,data=BT)  
abline(mymodel2)



Relationship is not really linear.

#Scatterplot of Eggs vs Fork Length Length  
plot(BT$FL,BT$Eggs,xlab="Total Length (mm)",ylab="No. eggs")  
#Linear regression for Eggs vs TL  
mymodel3=lm(Eggs~FL,data=BT)  
abline(mymodel3)



Relationship is not really linear.