Assignment 3

Paul Isaacson

### 1

### a.

c <- read.csv('Cars 2005.csv')  
attach(c)

## The following objects are masked from c (pos = 3):  
##   
## Cruise, Cylinder, Doors, Leather, Liter, Make, Mileage, Model,  
## Price, Sound, Trim, Type

## The following objects are masked from c (pos = 4):  
##   
## Cruise, Cylinder, Doors, Leather, Liter, Make, Mileage, Model,  
## Price, Sound, Trim, Type

## The following objects are masked from c (pos = 5):  
##   
## Cruise, Cylinder, Doors, Leather, Liter, Make, Mileage, Model,  
## Price, Sound, Trim, Type

## The following objects are masked from c (pos = 6):  
##   
## Cruise, Cylinder, Doors, Leather, Liter, Make, Mileage, Model,  
## Price, Sound, Trim, Type

## The following objects are masked from c (pos = 8):  
##   
## Cruise, Cylinder, Doors, Leather, Liter, Make, Mileage, Model,  
## Price, Sound, Trim, Type

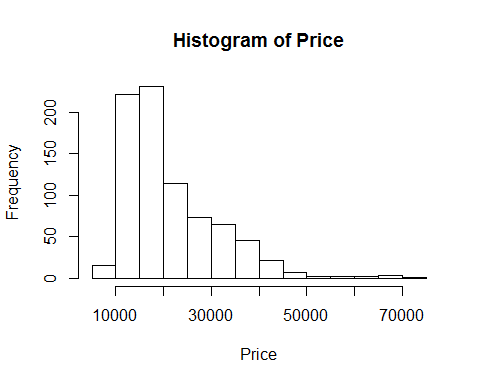
## The following objects are masked from c (pos = 9):  
##   
## Cruise, Cylinder, Doors, Leather, Liter, Make, Mileage, Model,  
## Price, Sound, Trim, Type

## The following objects are masked from c (pos = 10):  
##   
## Cruise, Cylinder, Doors, Leather, Liter, Make, Mileage, Model,  
## Price, Sound, Trim, Type

## The following objects are masked from c (pos = 11):  
##   
## Cruise, Cylinder, Doors, Leather, Liter, Make, Mileage, Model,  
## Price, Sound, Trim, Type

### b.

hist(params$cars$Price, xlab="Price", main="Histogram of Price")



The distribution is right-skewed.

### c.

criteria <- which((params$cars$Price > 10000) & (params$cars$Price <20000))  
length(criteria) / length(params$cars$Price)

## [1] 0.5621891

### d.

mean(params$cars$Price)

## [1] 21343.14

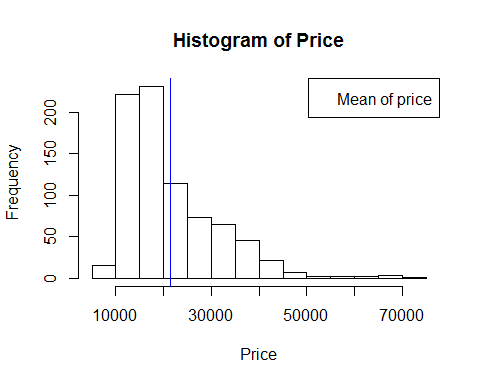
median(params$cars$Price)

## [1] 18025

The mean is larger. This makes sense because the distribution is right-skewed and the mean is more sensitive to outliers.

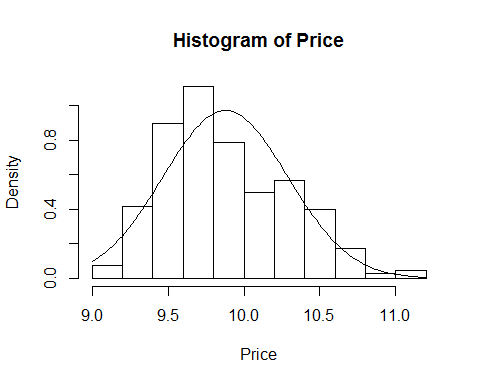
### e.

hist(params$cars$Price, xlab="Price", main="Histogram of Price")  
abline(v=mean(params$cars$Price), col="blue")  
legend("topright", legend=c("Mean of price"), col=c("blue"))



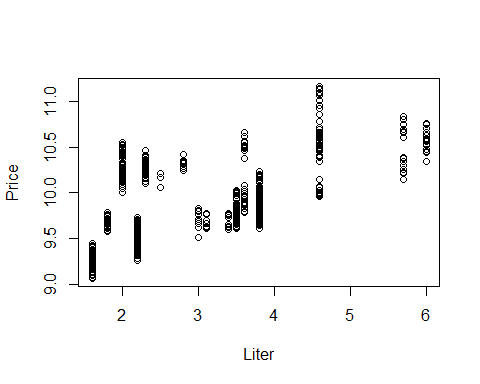
### f.

logPrice <- log(params$cars$Price)  
hist(log(params$cars$Price), prob=TRUE, xlab="Price", main="Histogram of Price")  
  
f <- function(x){dnorm(x, mean(log(params$cars$Price)), sd(log(params$cars$Price)))}  
curve(f(x), add=TRUE)



### g.

plot(log(params$cars$Price) ~ params$cars$Liter, xlab="Liter", ylab="Price")



The relationship between the two variables is weak. At each liter size, there is a wide range of transformed prices.

### h.

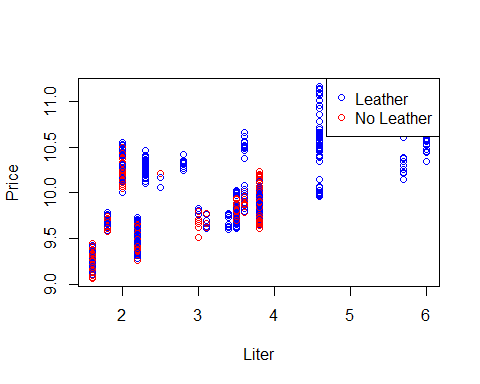
cor(params$cars$Liter, log(params$cars$Price))

## [1] 0.5904097

The correlation tells us there is a modest positive between the two variables.

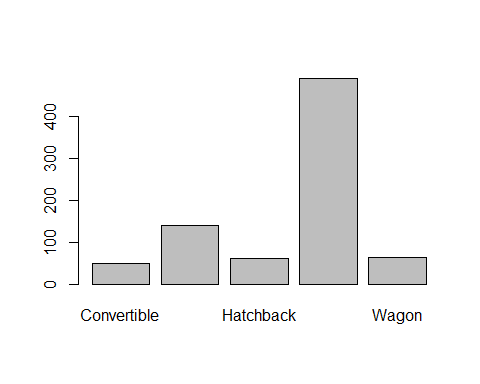
### i.

mycolors <- rep("red", length(params$cars$Price))  
criteria <- which(params$cars$Leather==TRUE)  
mycolors [criteria] <- "blue"  
plot(log(params$cars$Price) ~ params$cars$Liter, col=mycolors, xlab="Liter", ylab="Price")  
legend("topright", legend=c("Leather", "No Leather"), col=c("blue", "red"), pch=21)



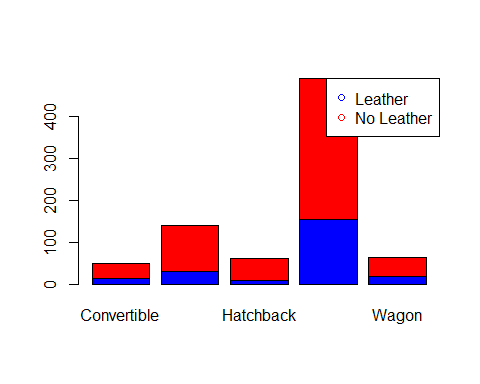
### j.

counts <- table(params$cars$Type)  
barplot(counts)



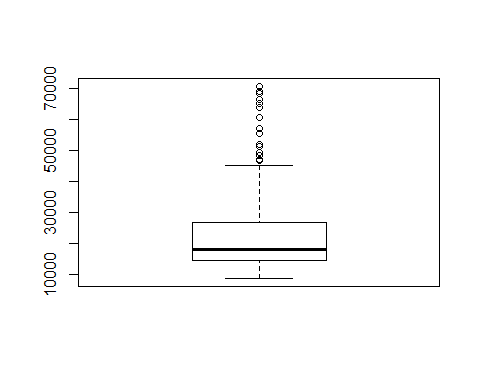
### k.

counts <- table(params$cars$Leather, params$cars$Type)  
barplot(counts, col=c("blue","red"))  
legend("topright", legend=c("Leather", "No Leather"), col=c("blue", "red"), pch=21)



### l.

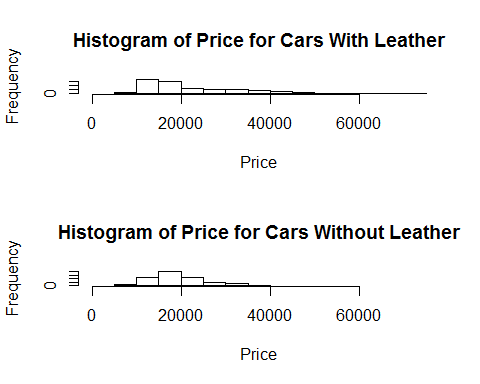
boxplot(params$cars$Price)



The IQR is fairly tight and on the lower end of the range. As such, most car prices are between approximately 15,000 and 25,000. Also, cars above 45,000 are considered outliers. Meaning, those are luxury cars not common for the average consumer.

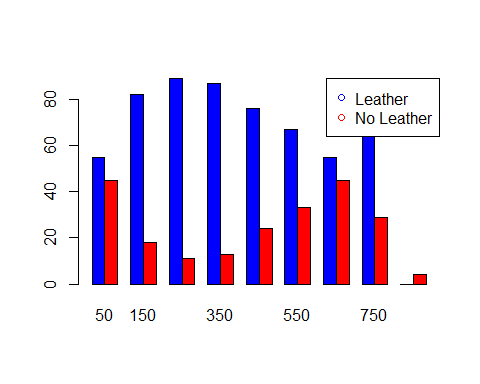
### m.

par(mfrow = c(2,1))  
hasLeather = which(params$cars$Leather==TRUE)  
hasNoLeather = which(params$cars$Leather==FALSE)  
hist(params$cars$Price[hasLeather], xlim=c(0,75000), xlab="Price", main="Histogram of Price for Cars With Leather")  
hist(params$cars$Price[hasNoLeather], xlim=c(0,75000), xlab="Price", main="Histogram of Price for Cars Without Leather")



### n.

library(plotrix)  
hasLeather = which(params$cars$Leather==TRUE)  
hasNoLeather = which(params$cars$Leather==FALSE)  
multhist(list(hasLeather, hasNoLeather), col=c("blue", "red"))  
legend("topright", legend=c("Leather", "No Leather"), col=c("blue", "red"), pch=21)



### 2.

### a.

data(Mammals, package="quantreg")  
head(Mammals)

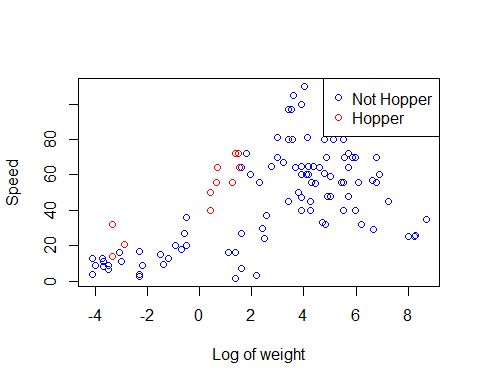
## weight speed hoppers specials  
## 1 6000 35 FALSE FALSE  
## 2 4000 26 FALSE FALSE  
## 3 3000 25 FALSE FALSE  
## 4 1400 45 FALSE FALSE  
## 5 400 70 FALSE FALSE  
## 6 350 70 FALSE FALSE

### b.

I decided to transform weight. Otherwise, the weights were packed too closely on the lower end. The median weight was 34 and the mean weight 278.6882. The variable is right-skewed.

### c.

data(Mammals, package="quantreg")  
mycolors <- rep("blue", length(Mammals$weight))  
criteria <- which(Mammals$hopper==TRUE)  
mycolors[criteria] <- "red"  
plot(log(Mammals$weight),Mammals$speed, col=mycolors, xlab="Log of weight", ylab="Speed")  
legend("topright", legend=c("Not Hopper", "Hopper"), col=c("blue", "red"), pch=21)



cor(log(Mammals$weight), Mammals$speed)

## [1] 0.5751193

There are only 11 hoppers in the group. But, when compared to animals of similar weight, they do appear to be the fastest. I chose the scatterplot because we are looking at the relatonship between two variables. Also I chose to color the hoppers a different color so they easily stand out. Finally, I added the correlation to show the strength of the relationship. In this case, there is a slight positive correlation.