1. Because all normal distributions have their data similarly spread out, we can look at the case for the standard normal and find the percentage of outliers in this way. The variable lower_outside corresponds to the boundary of the lower outside bar. Then I find the probability that the data lies below this. Since the distribution is symmetric, I multiply by 2 to find the total probability that the data lies either below the lower outside bar or above the upper outside bar. **I expect about .7% of the data to lie outside the outside bars.**

```
> lower_outside = qnorm(.25) - 1.5*(qnorm(.75) - qnorm(.25))
> 2*pnorm(lower_outside)
[1] 0.006976603
```

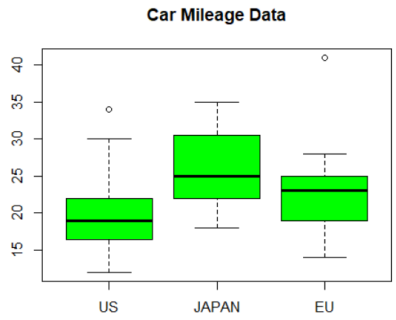
2. (1) (CODE BELOW) The five number summaries of the mileages are:

US: 12.0 16.5 19.0 22.0 34.0

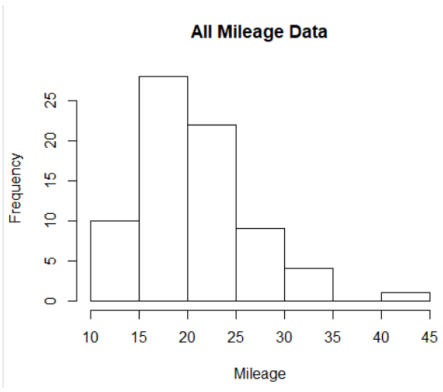
Japan: 18.0 22.0 25.0 30.5 35.0

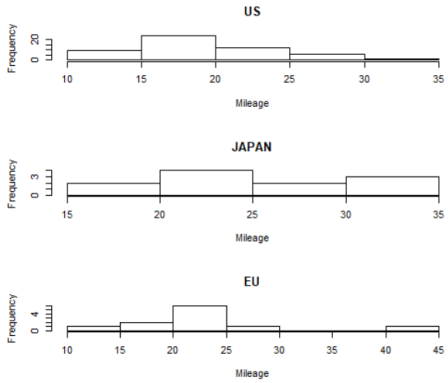
Europe: 14 19 23 25 41

2. (2) (CODE BELOW) From the boxplots, I can conclude that **Japan's cars have the best fuel efficiency, and next Europe's cars, and last US's cars**. This conclusion is based on the positioning of the middle 50% (interquartile range) of the data of each group, as well as the median. I can also conclude by looking at the outside bars, that US's mileages are the most spread out with 1 outlier, while Europe's mileages are the least spread out with 1 very high outlier.



2. (3) (CODE BELOW) From the histograms, I can conclude about the same thing as from the boxplots: **order of best fuel efficiency is Japan, Europe, and last, US.** However, with the histograms, it is harder to identify which bars correspond to outside bars and outliers, and which bars have data that are part of the interquartile range. Also, it seems as if Japan has the least spread, but the most evenly distributed IQR, while US and Europe have an obvious peak.





HW1.R

Source on Save

```r
#get car data
CarData <- read.table("carc.dat")
#extract mileage and headqtr from table
x <- CarData[c(3,14)]

#PART 1: fivenum summaries of US, Japan, EU
fivenum(x[x$V14 == 1, 1])
fivenum(x[x$V14 == 2, 1])
fivenum(x[x$V14 == 3, 1])

#PART 2: factor headqtrs and draw boxplot
headquarter <- factor(x[,2])
levels(headquarter) <- c("US","JAPAN", "EU")
x <- cbind(x, headquarter)
colnames(x)=c("Mileage", "HeadInt", "HeadFac")
boxplot(Mileage ~ HeadFac, data=x, at=1:3, main="Car Mileage Data", col='green')

#PART 3: draw histogram for total mileage & country groups
hist(x[,1], main="All Mileage Data", xlab="Mileage")

par(mfrow = c(3, 1))
hist(x[headquarter=="US",1], main="US", xlab="Mileage")
hist(x[headquarter=="JAPAN",1], main="JAPAN", xlab="Mileage")
hist(x[headquarter=="EU",1], main="EU", xlab="Mileage")
```