Student - Assignment

Student

9/26/2022

Question 1

```
data1 <- read.csv("mlbstats.csv", header = T)
stem(data1$HR)
```

```
##
##   The decimal point is 1 digit(s) to the right of the |
##
##   0 | 001333344
##   0 | 5666677778888889999
##   1 | 00000111111122222233333333344
##   1 | 55555555555666667777888899999999
##   2 | 00011111111223344444
##   2 | 55666777778889
##   3 | 012333
##   3 | 556
##   4 |
##   4 |
##   5 |
##   5 | 5
```

```
print("The HR distribution doesn't seem to be symmetric, but rightly/positively skewed.", quote = F)
```

```
## [1] The HR distribution doesn't seem to be symmetric, but rightly/positively skewed.
```

```
which(data1$HR == 55)
```

```
## [1] 7
```

```
print("The value 55 can be considered as an outlier.
Using the function which, it can be found,
      that 55 is located in the 7th position.", quote = F)
```

```
## [1] The value 55 can be considered as an outlier. \nUsing the function which, it can be found, \n
```

```
mean(data1$HR)
```

```
## [1] 16.62595
```

```
sd(data1$HR)
```

```
## [1] 8.898
```

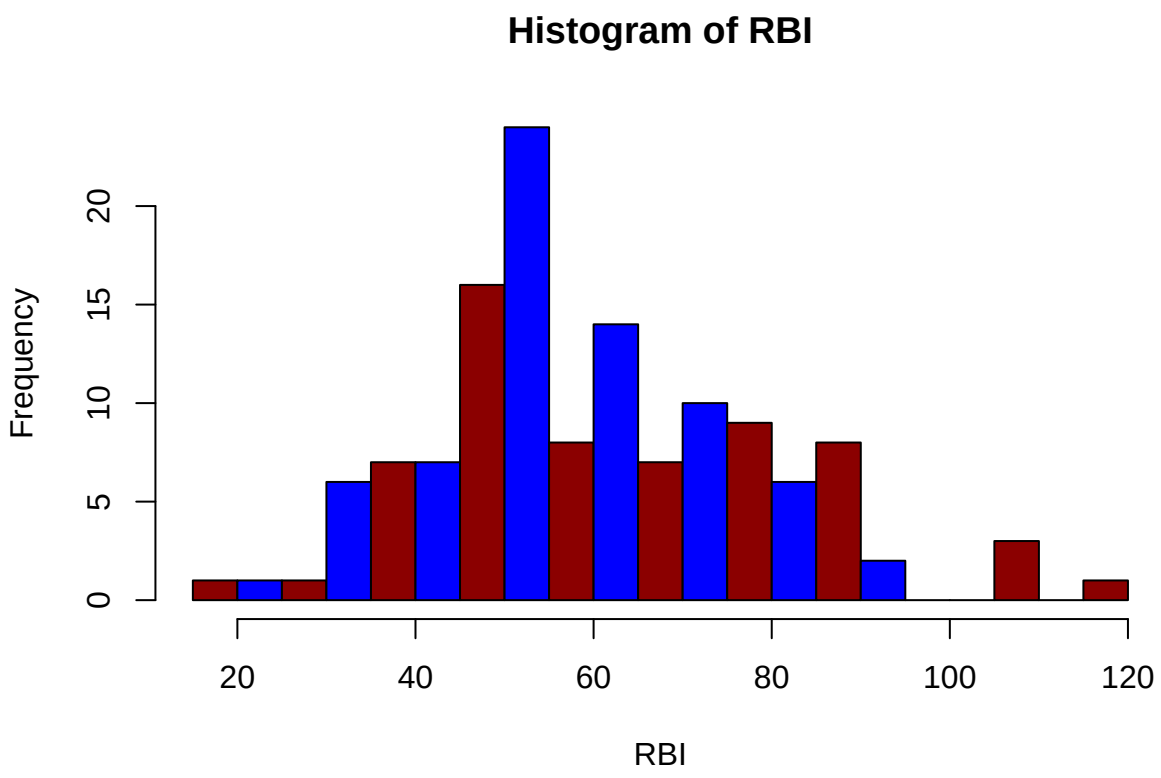
```
quantile(data1$HR, 0.25)
```

```
##  25%
## 10.5
```

```
quantile(data1$HR, 0.75)
```

```
##  75%
## 21.5
```

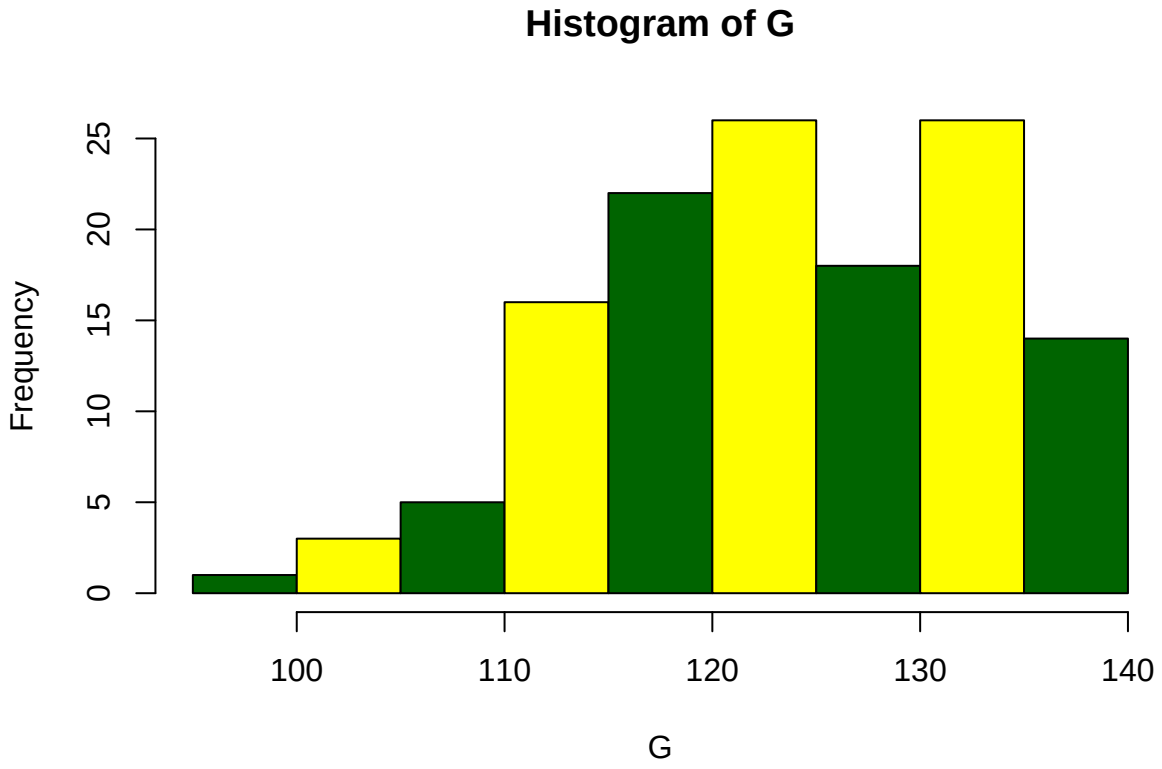
```
hist(data1$RBI, main = "Histogram of RBI", breaks = 20, col = c("darkred","blue"), xlab = "RBI")
```



```
print("The variable RBI is a rather symmetric distribution.", quote = F)
```

```
## [1] The variable RBI is a rather symmetric distribution.
```

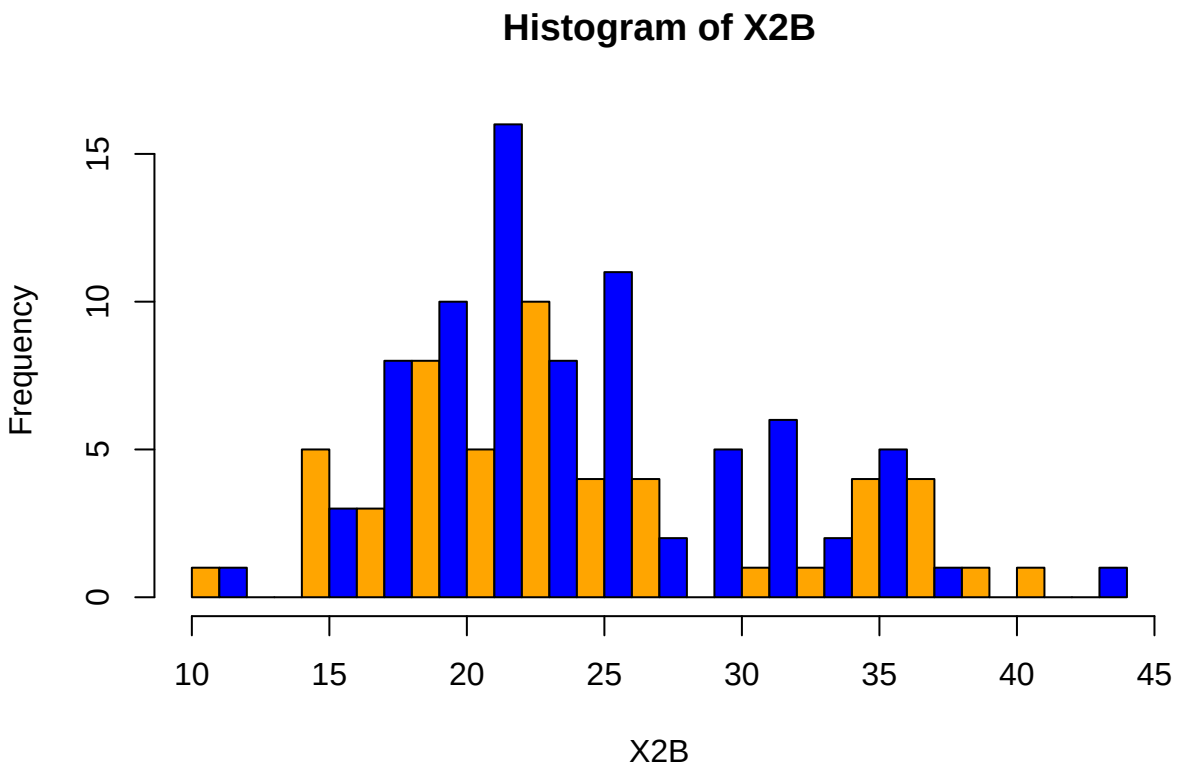
```
hist(data1$G, main = "Histogram of G", breaks = 8, col = c("darkgreen","yellow"), xlab = "G")
```




```
print("The variable G is a rather left/negatively skewed distribution.", quote = F)
```

```
## [1] The variable G is a rather left/negatively skewed distribution.
```

```
hist(data1$X2B, main = "Histogram of X2B", breaks = 25, col = c("orange","blue"), xlab = "X2B")
```




```
print("The variable X2B is a rather symmetric distribution.", quote = F)
```

```
## [1] The variable X2B is a rather symmetric distribution.
```

Question 2

```
data2 <- read.csv("cellservice.csv", header = T)
data2_ATT <- subset(data2, select = c(ï..City,AT.T))
data2_Sprint <- subset(data2, select = c(ï..City,Sprint))
data2_T.Mobile <- subset(data2, select = c(ï..City,T.Mobile))
data2_Verizon <- subset(data2, select = c(ï..City,Verizon))

summary(data2_ATT$AT.T)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   66.00   68.00   71.00   70.65   73.00   75.00
```

```
summary(data2_Sprint$Sprint)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   63.00   65.00   66.00   66.10   67.25   69.00
```

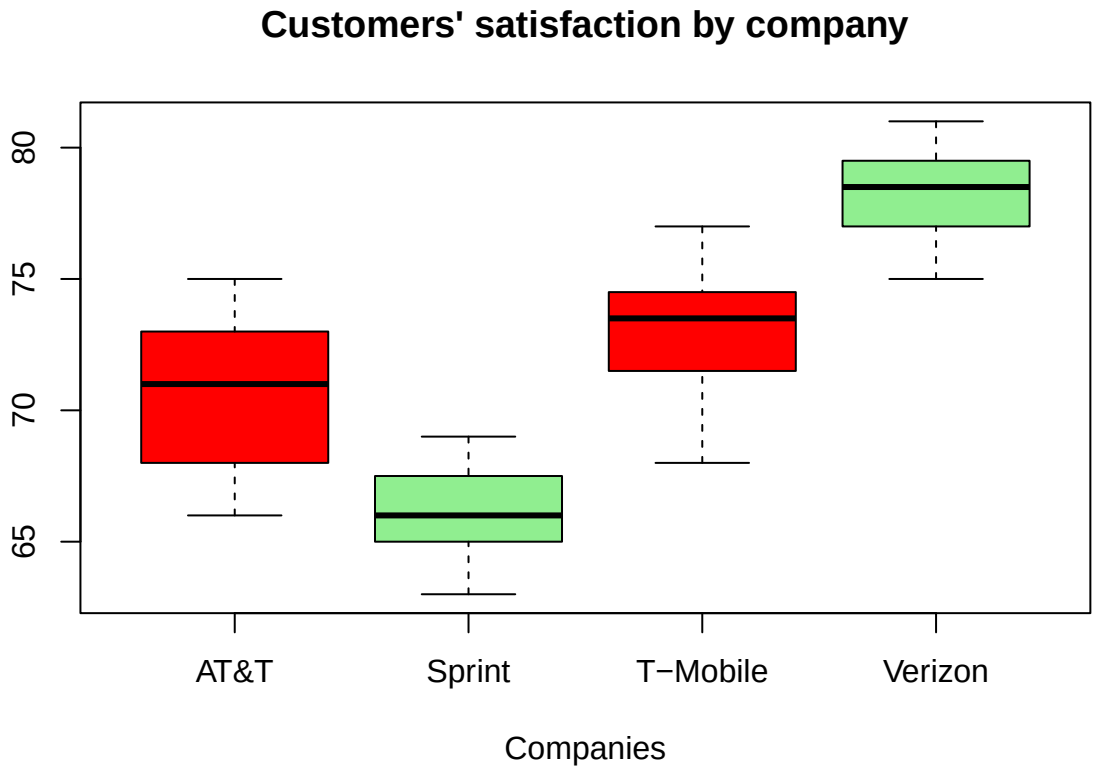
```
summary(data2_T.Mobile$T.Mobile)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   68.00   71.75   73.50   73.15   74.25   77.00
```

```
summary(data2_Verizon$Verizon)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   75.00   77.00   78.50   78.35   79.25   81.00
```

```
boxplot(data2$AT.T, data2$Sprint, data2$T.Mobile, data2$Verizon,
        xlab = "Companies", col = c("red","light green"),
        names = c("AT&T","Sprint","T-Mobile","Verizon"),
        main = "Customers' satisfaction by company")
```

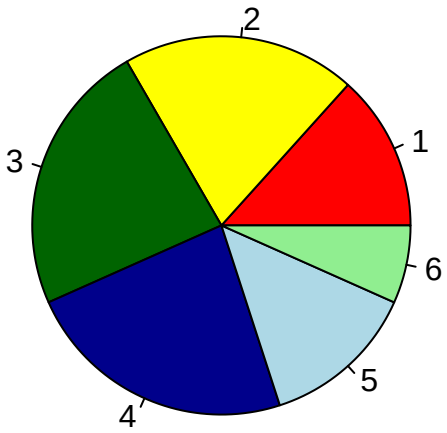


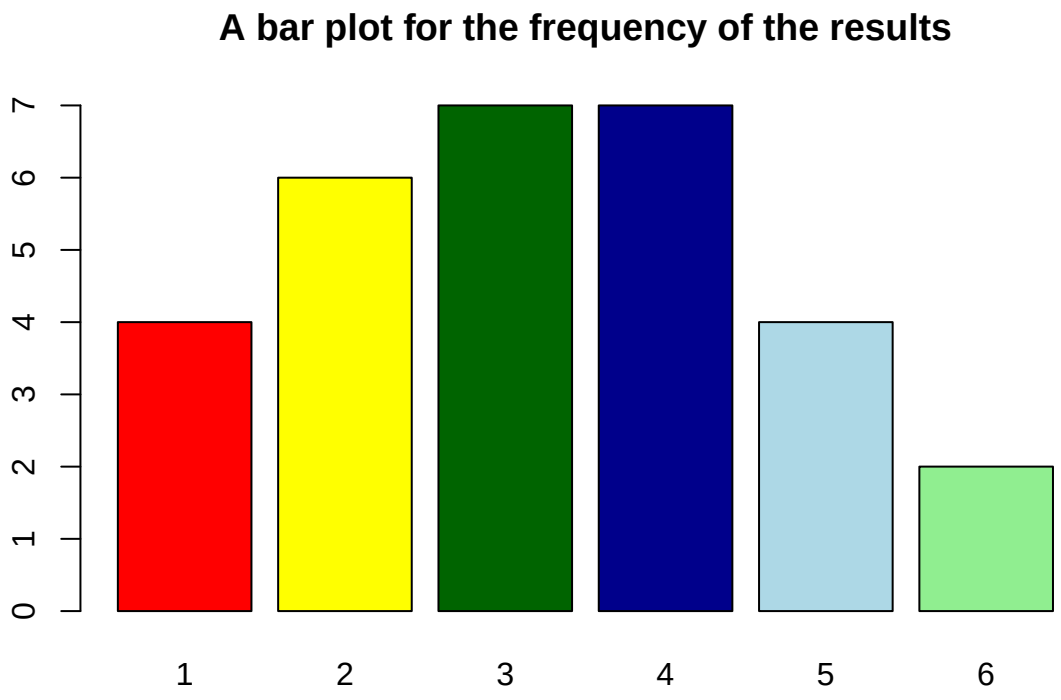
As for the boxplots:

According to the boxplots, the company with the highest rates of customers' satisfaction is Verizon, than T-mobile, than AT&T and the last company, with the lowest rates of customers' satisfaction, is Sprint. According to these results, the most recommended company is Verizon. It seems that Verizon's rating is greater than all the others, and far greater than Sprint's rating, which is the lowest.

Question 3

```
## data3
## 1 2 3 4 5 6
## 4 6 7 7 4 2
```






```
## [1] Number of students that are taking more than 3 courses:
```

```
## [1] 13
```

Question 4

```
data4 <- seq(from = 2, to = 42, by = 4)
data4
```

```
##  [1]  2  6 10 14 18 22 26 30 34 38 42
```

```
data4_log <- log(data4)
data4_log
```

```
##  [1] 0.6931472 1.7917595 2.3025851 2.6390573 2.8903718 3.0910425 3.2580965
##  [8] 3.4011974 3.5263605 3.6375862 3.7376696
```

```
data4_log_no_3_to_7 <- data4_log[-c(3,4,5,6,7)]
data4_log_no_3_to_7
```

```
## [1] 0.6931472 1.7917595 3.4011974 3.5263605 3.6375862 3.7376696
```

```
length_c <- length(data4_log_no_3_to_7)
length_c
```

```
## [1] 6
```

```
data4_log_no_3_to_7_sorted <- sort(data4_log_no_3_to_7)
data4_log_no_3_to_7_sorted
```

```
## [1] 0.6931472 1.7917595 3.4011974 3.5263605 3.6375862 3.7376696
```

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.