STAT 260 R Assignment 1

Rodriguez Castro Raul V01030827

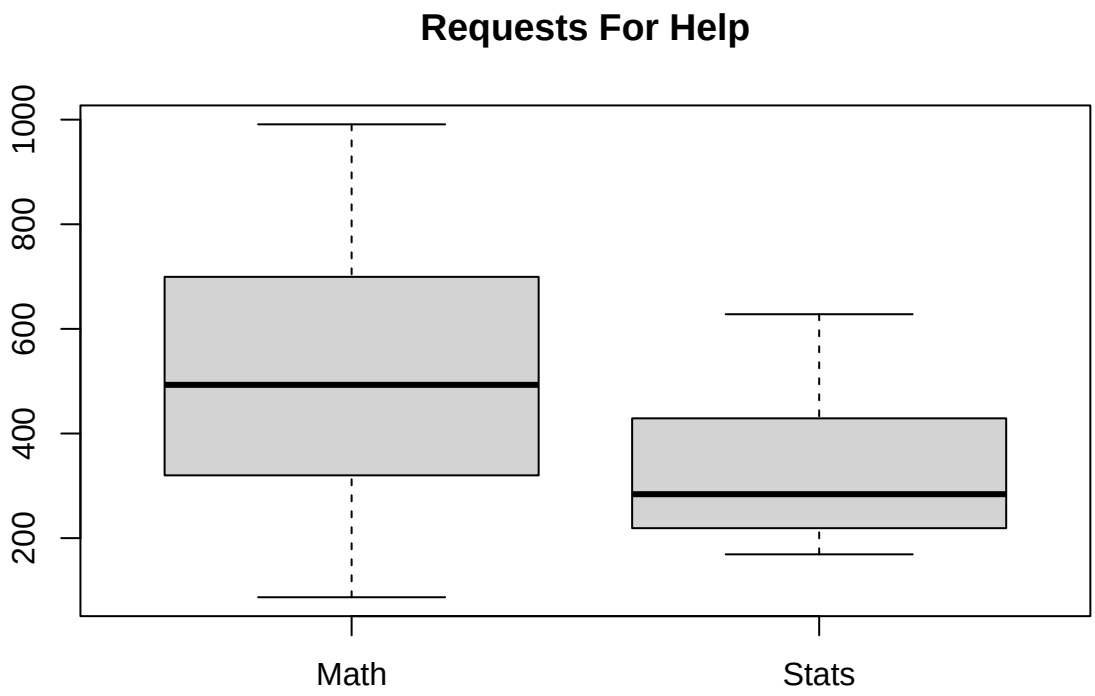
Question 1

a), b)

```
requests.help.stats = scan(
  text = "578 183 423 429 188 295 273 232 169 271 251 219 313 571 467 628 402 175",
  what = numeric()
)

requests.help.math = scan(
  text = "544 87 743 391 676 724 991 801 498 175 392 108 402 305 723 662 599 488 319 321",
  what = numeric()
)

boxplot(
  requests.help.math, requests.help.stats,
  names = c("Math", "Stats"),
  main = "Requests For Help"
  )
```



c)

Mean For Math

```
mean(requests.help.math)
```

```
## [1] 497.45
```

Standard Deviation For Math

```
sd(requests.help.math)
```

```
## [1] 243.3802
```

Mean For Stats

```
mean(requests.help.stats)
```

```
## [1] 337.0556
```

Standard Deviation For Stats

```
sd(requests.help.stats)
```

```
## [1] 149.12
```

d)

Based on the means of the two data sets, Math seems to be harder since there are more requests for help than for Stats. The Standard Deviation also suggests (assuming that the data set contains weekly requests throughout a term) that the Math students requested help more evenly throughout the term. While The Stats students show greater variability meaning the weren't as consistent as Math students asking for help.

Question 2

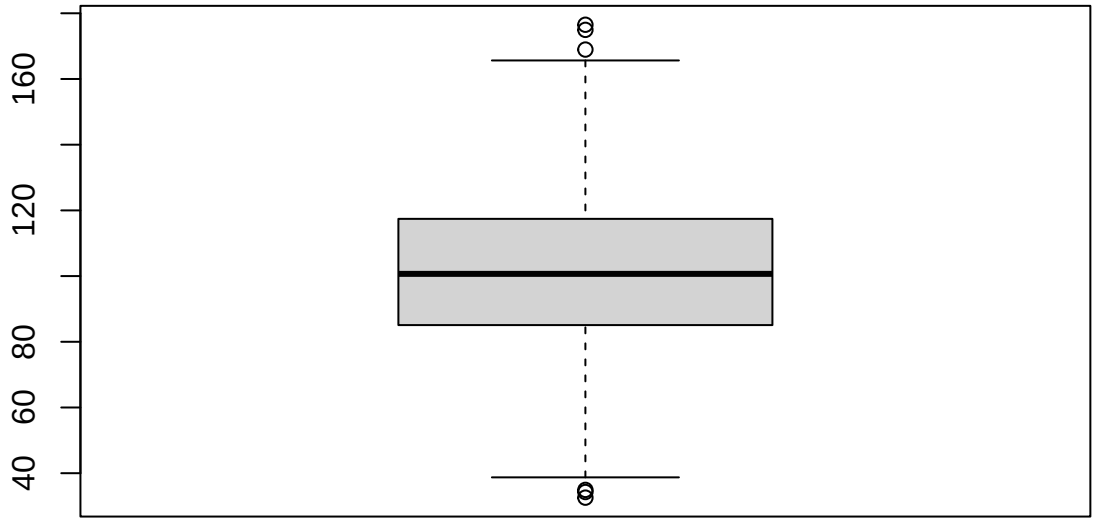
a)

```
set.seed(14253)
simvector = rnorm(600, 100, 24)
summary(simvector)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   32.56   85.15  100.67  101.22  117.38  176.51
```

```
hist(simvector)
```

```
boxplot(simvector)
```



b)

Interquartile Range

```
q1 = quantile(simvector, 0.25)
q3 = quantile(simvector, 0.75)

iqr = q3 - q1
iqr
```

```
##      75%
## 32.22882
```

c)

Lower and Upper Limits to identify outliers

```
#q1 = quantile(simvector, 0.25)
#q3 = quantile(simvector, 0.75)

#iqr = q3 - q1
```

```
lower.limit = q1 - (1.5 * iqr)
upper.limit = q3 + (1.5 * iqr)

lower.limit
```

```
##      25%
## 36.80395
```

```
upper.limit
```

```
##      75%
## 165.7192
```

d)

The histogram displays a symmetric distribution with bimodal characteristics. It has two prominent peaks on the left and right side of the histogram's center. There is a rightward tail that is slightly noticeable indicating skewness towards higher values.

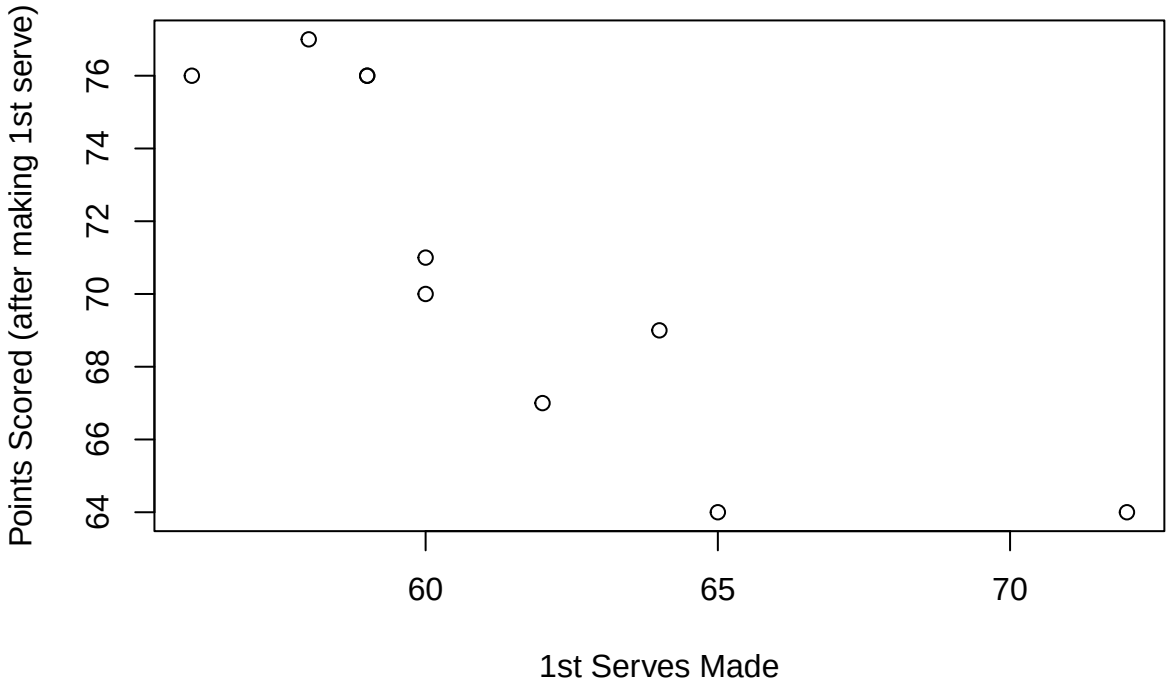
The boxplot shows two center lines indicating a bimodal distribution. There are notably outliers present above the upper bounds and below the lower bounds of the box. indicating that there might be some data points that deviate significantly from the rest.

Question 3

a)

```
x = as.numeric(strsplit("56, 58, 59, 59, 60, 60, 62, 64, 65, 72", ",")[[1]])
y = as.numeric(strsplit("76, 77, 76, 76, 71, 70, 67, 69, 64, 64", ",")[[1]])

plot(x, y, xlab = "1st Serves Made", ylab = "Points Scored (after making 1st serve)", main = "")
```



b)

Correlation Coefficient:

```
corr_coeff = cor(x,y)
corr_coeff
```

```
## [1] -0.850068
```

c)

It is reasonable to model this data as a linear relationship because the correlation coefficient (-0.85) is close to -1. Furthermore, there is a negative slope indicating a negative relationship between the first serve and the points scored meaning the more often a proffesional player makes the 1st serve the less overall points they score afterward.