# R Assignment 1 Suggested Solutions

## Question 1

### 1a

Cal 1 Scores:

66 65 80 66 73 94 100 93 79 89 64 90 63 88 85 67 77 82 82 39 100 82 60 56 66 72 74 67 87 61 50 58 66 100 93 92 60 74 59 94

Cal 2 Scores: 74 77 64 63 59 69 72 57 64 55 69 40 60 54 80 45 58 60 65 71 60 61 66 74 68 71 73 99 66 63 69 69 75 47 50

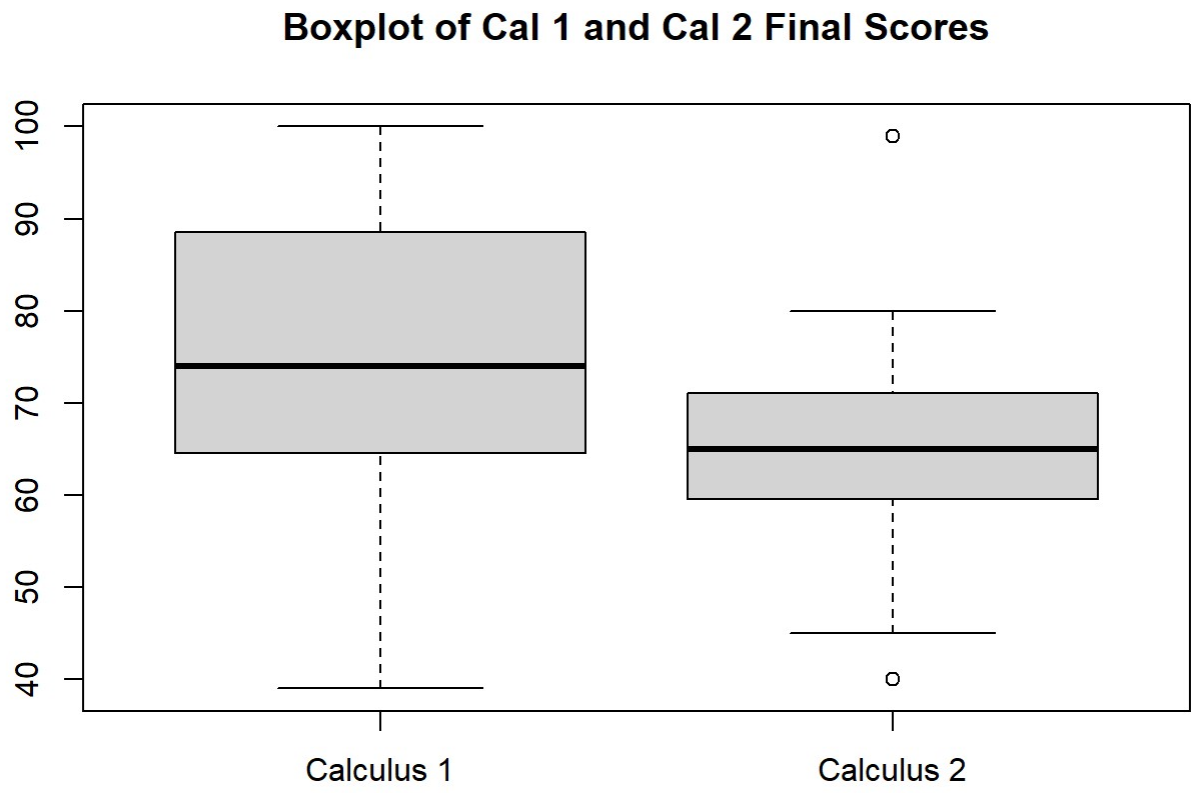
You can also use the scan() and copy and paste data

```
cal1 <- c(66,65,80,66,73,94,100,93,79,89,64,90,63,88,85,67,77,82,82,39,100,82,60,56,66,72,74,
67,87,61,50,58,66,100,93,92,60,74,59,94)

cal2 <- c(74,77,64,63,59,69,72,57,64,55,69,40,60,54,80,45,58,60,65,71,60,61,66,74,68,71,73,9
9,66,63,69,69,75,47,50)


boxplot(cal1, cal2, names = c("Calculus 1", "Calculus 2"),main = "Boxplot of Cal 1 and Cal 2
Final Scores")
```

1b




1c

```
mean(cal1)
```

```
## [1] 75.325
```

```
sd(cal1)
```

```
## [1] 15.07313
```

```
mean(cal2)
```

```
## [1] 64.77143
```

```
sd (cal2)
```

```
## [1] 10.9681
```

## 1d

Suggested summary statistics

```
summary(cal1)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   39.00   64.75   74.00   75.33   88.25  100.00
```

```
summary(cal2)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   40.00   59.50   65.00   64.77   71.00   99.00
```

```
#coefficient of variation

sd(cal1)/mean(cal1)
```

```
## [1] 0.200108
```

```
sd(cal2)/mean(cal2)
```

```
## [1] 0.1693354
```

Based on the boxplot and the summary statistics, it appears that overall Cal 1 had a higher final score than that of Cal 2. The mean and median is higher in Cal 1. The scores for the Cal 1 were more spread out than that of Cal 2, evidenced by the larger coefficient of variation and the inter-quartile range.

# Question 2

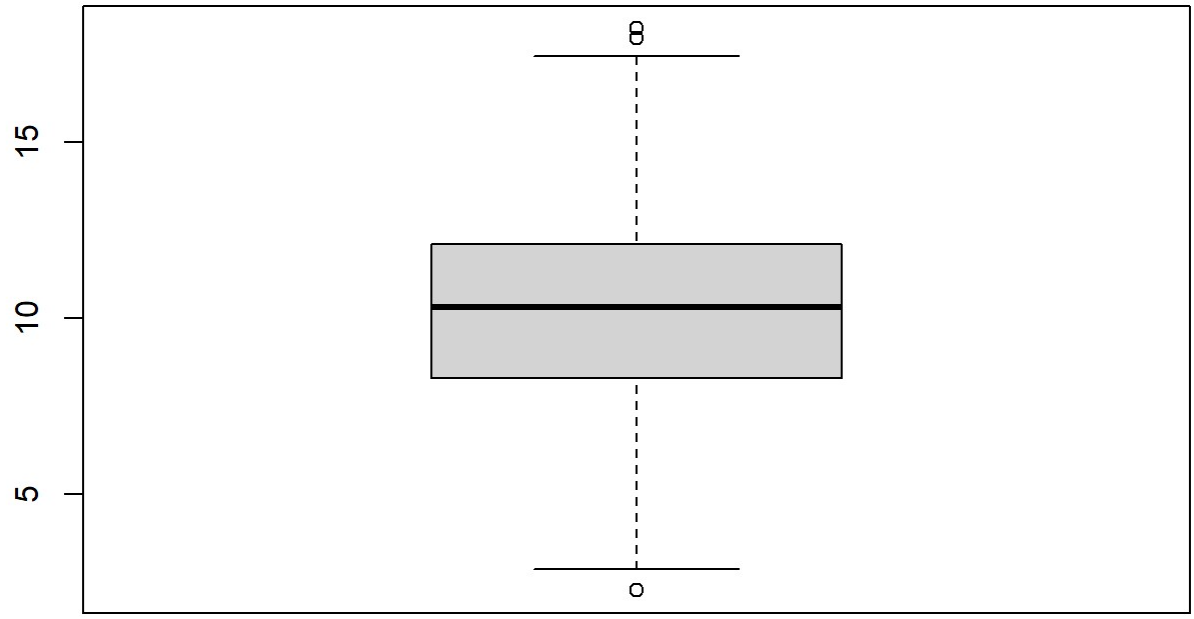
```
set.seed(12345)
simvector <- rnorm(500, 10, 3)
# The above command generates a sample of 500 data points from a normal distribution, which i
s bell-shaped
# with a single peak and symmetric around 10.
summary(simvector)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   2.254   8.285  10.307  10.247  12.087  18.242
```

## Part a

```
hist(simvector)
```


```
boxplot(simvector)
```




## Part b

```
iqr <- 12.087 - 8.285
iqr
```

```
## [1] 3.802
```

```
#or
IQR(simvector)
```

```
## [1] 3.802481
```

## Part c

```
l <- 8.285 - 1.5*iqr
u <- 12.087 + 1.5*iqr

l
```

```
## [1] 2.582
```

```
u
```

```
## [1] 17.79
```

## Part d (Suggested)

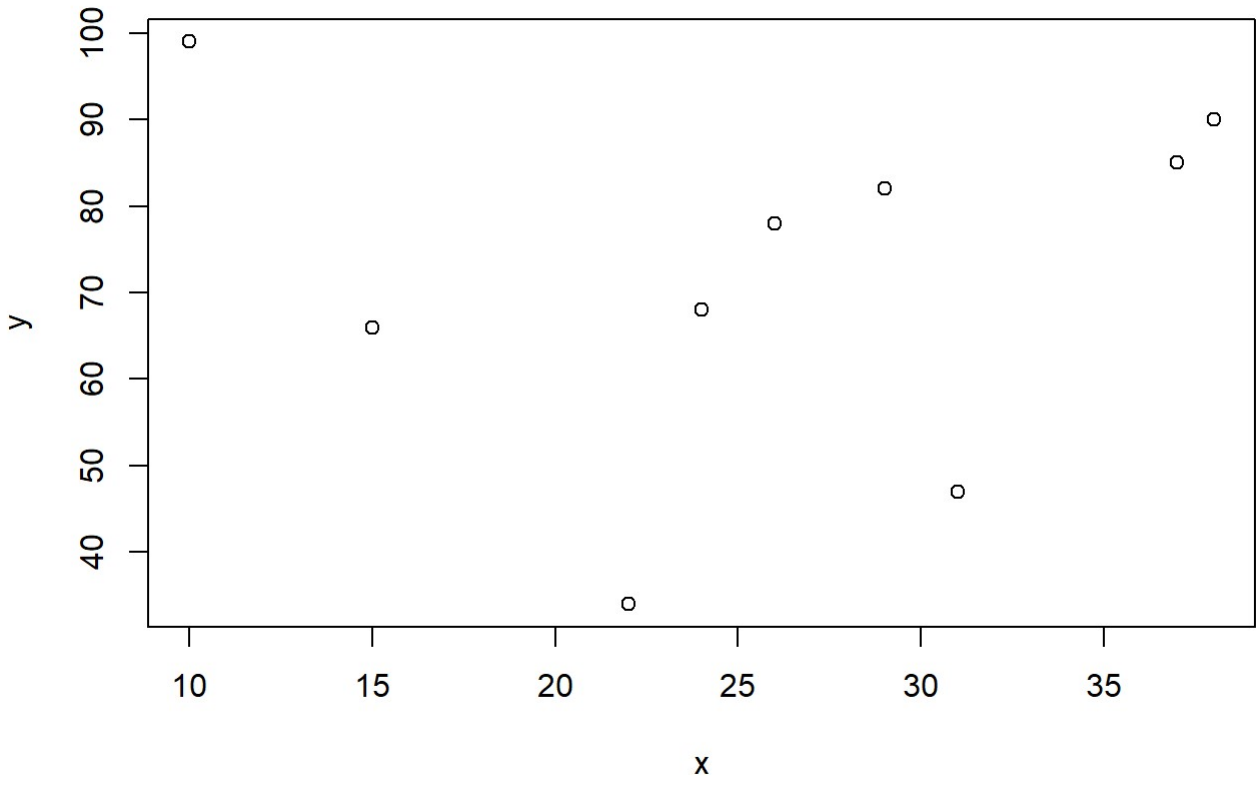
From the histogram and boxplot, the data has one peak (unimodal) and roughly symmetric.

# Question 3

## Part a

```
x <- c(29, 15, 26, 22, 31, 37, 38, 10, 24)
y <- c(82, 66, 78, 34, 47, 85, 90, 99, 68)

plot(x,y)
```




## Part b

```
cor(x,y)
```

```
## [1] 0.03742851
```

```
#You can use the formula and do it by hand.
```

There is a (very weak) positive association (or relation) between the amount of study time (x) and the Physical class score (y).

# Part c

The question has many correct answers. We will accept any comments that make sense.

You should NOT say that the class scores affect amount of study time since that there is the order of time. We do not accept the concept of time travel (at the moment).

Although it makes sense that the more time one spends on studying (one week before the test), the higher score they will achieve in the test, it may not be the only cause. There are other factor(s) which can also affect their test score, for example, how much time they spend time on studying during the course.