Assignment 2 Q1

Guru Shyaam Shankar

24/10/2021

```
#Question 1:


Hospital_Time <- read.csv("/Users/shyaamshankar/Downloads/Hospital_time.csv", header = TRUE)

#Now we need to filter the csv file so we can read the csv files individually for
#every Hospital,  to compare their Waiting Times. We know that the sample size is 14
#for every Hospital, and is represented by the nrow function.


Hospital_A <- dplyr::filter(Hospital_Time, Hospital=="Hospital A")
print(Hospital_A)
```

```
##    Waiting.time   Hospital
## 1            67 Hospital A
## 2            45 Hospital A
## 3            15 Hospital A
## 4            60 Hospital A
## 5            70 Hospital A
## 6            75 Hospital A
## 7            36 Hospital A
## 8            95 Hospital A
## 9            95 Hospital A
## 10          135 Hospital A
## 11           35 Hospital A
## 12           14 Hospital A
## 13           68 Hospital A
## 14           47 Hospital A
```

```
sd(Hospital_A$Waiting.time)
```

```
## [1] 33.08306
```

```
mean(Hospital_A$Waiting.time)
```

```
## [1] 61.21429
```

```
nrow(Hospital_A)
```

```
## [1] 14
```

```
Hospital_B <- dplyr::filter(Hospital_Time, Hospital=="Hospital B")
print(Hospital_B)
```

```
##    Waiting.time   Hospital
```

```
## 1            64 Hospital B
## 2            44 Hospital B
## 3            19 Hospital B
## 4            37 Hospital B
## 5            15 Hospital B
## 6            30 Hospital B
## 7            22 Hospital B
## 8            16 Hospital B
## 9            60 Hospital B
## 10           52 Hospital B
## 11           55 Hospital B
## 12           18 Hospital B
## 13           46 Hospital B
## 14           20 Hospital B
```

```
sd(Hospital_B$Waiting.time)
```

```
## [1] 17.7188
```

```
mean(Hospital_B$Waiting.time)
```

```
## [1] 35.57143
```

```
nrow(Hospital_B)
```

```
## [1] 14
```

```
Hospital_C <- dplyr::filter(Hospital_Time, Hospital=="Hospital C")
print(Hospital_C)
```

```
##    Waiting.time   Hospital
## 1            36 Hospital C
## 2            68 Hospital C
## 3            15 Hospital C
## 4            20 Hospital C
## 5            34 Hospital C
## 6            38 Hospital C
## 7            32 Hospital C
## 8            18 Hospital C
## 9            60 Hospital C
## 10           24 Hospital C
## 11           32 Hospital C
## 12           18 Hospital C
## 13           20 Hospital C
## 14           23 Hospital C
```

```
sd(Hospital_C$Waiting.time)
```

```
## [1] 15.794
```

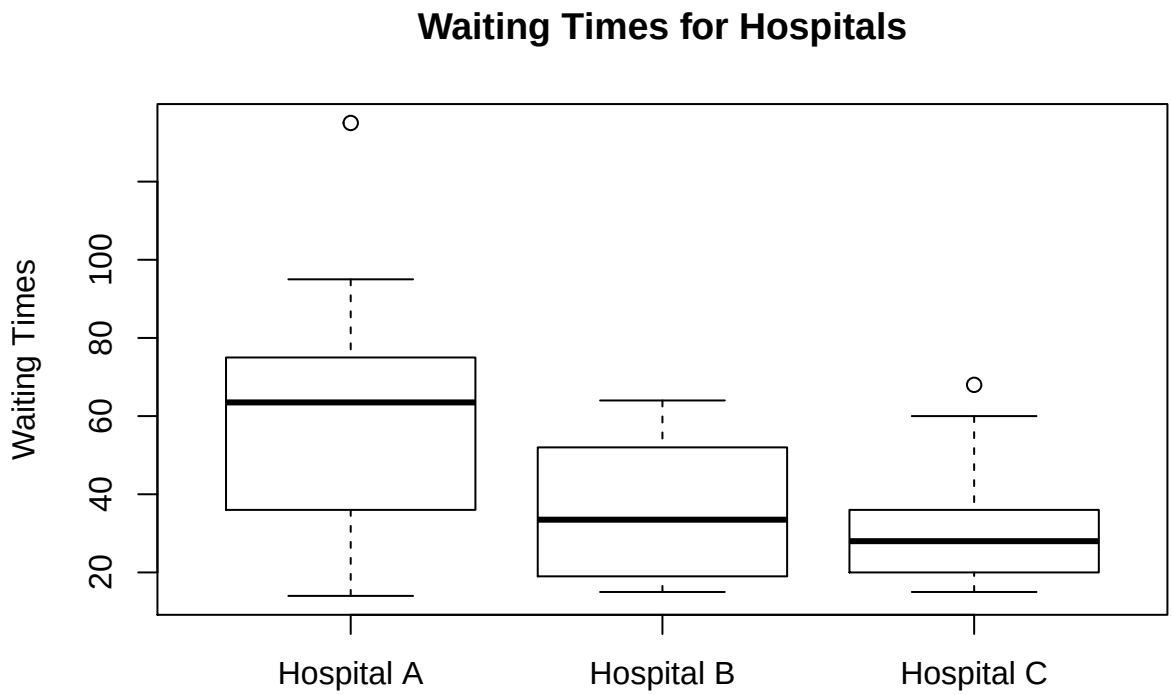
```
mean(Hospital_C$Waiting.time)
```

```
## [1] 31.28571
```

```
nrow(Hospital_C)
```

```
## [1] 14
```

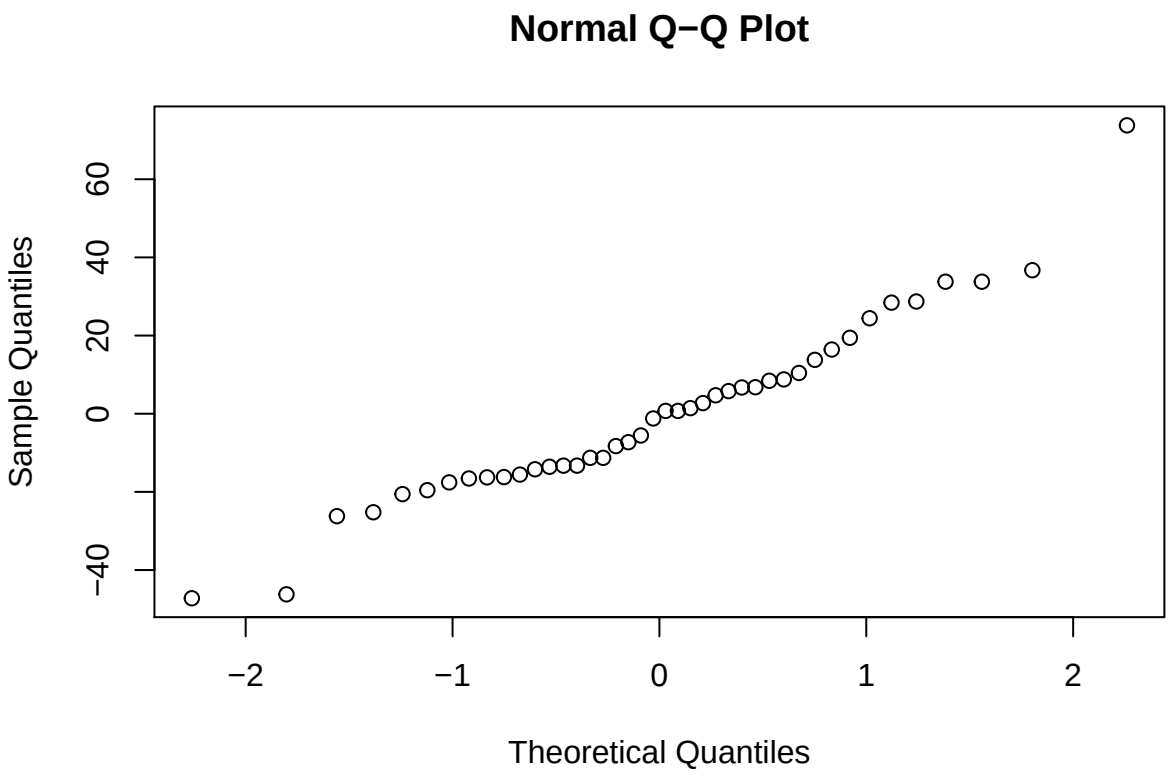
```r
#Now we plot the boxplot for the Waiting Times and their respective Hospitals.

boxplot(Waiting.time ~ Hospital, data = Hospital_Time,
        ylab = "Waiting Times", main = "Waiting Times for Hospitals")
```



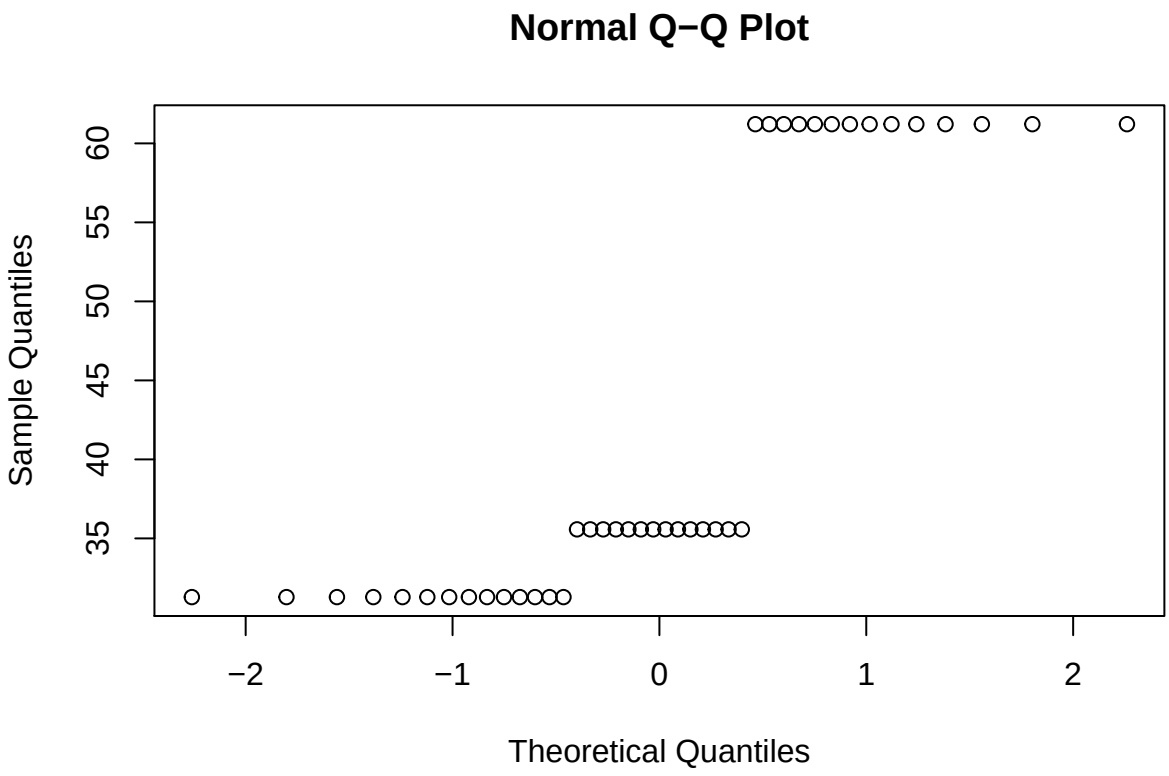

```r
#From the boxplot we can see that the data is suitable for an ANOVA test, as it tests
#response to a treatment.

Hospital_Time.aov <- aov(Waiting.time ~ Hospital, data = Hospital_Time)
anova(Hospital_Time.aov)
```

```
## Analysis of Variance Table
##
## Response: Waiting.time
##           Df  Sum Sq Mean Sq F value   Pr(>F)
## Hospital   2  7334.3  3667.2  6.6358 0.003308 **
## Residuals 39 21552.6   552.6
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

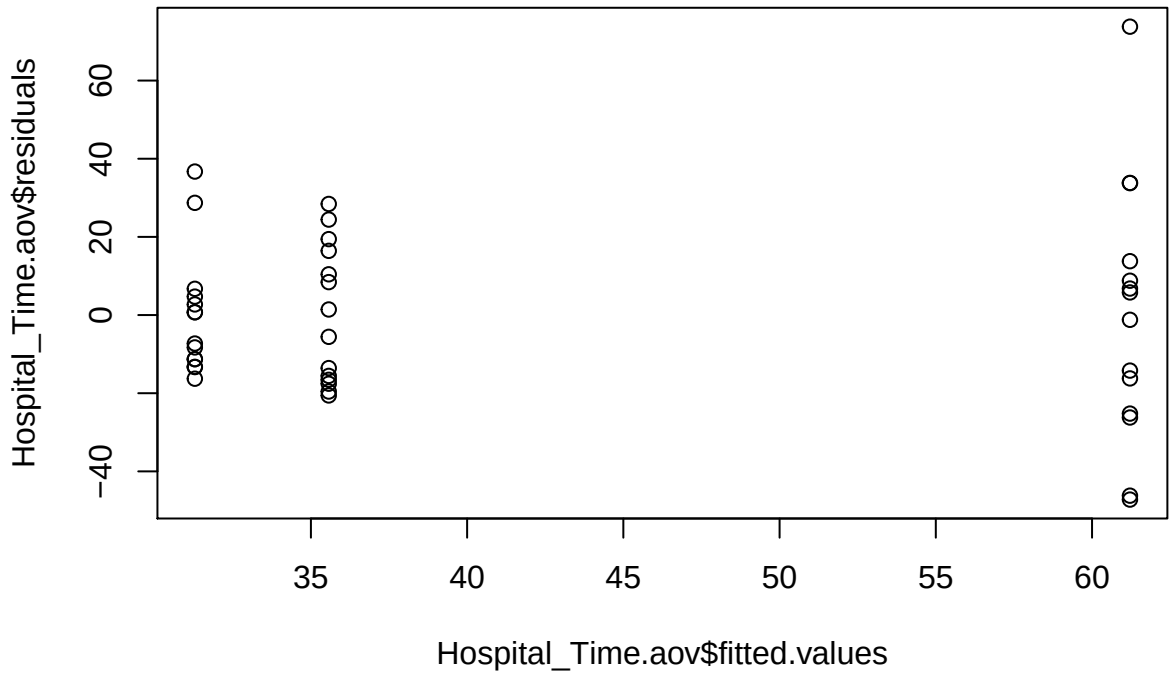
```r
#Therefore since P-Value < 0.05, we reject HO, that is to say we have sufficient
#evidence to state that one of the means is different to the others.

qqnorm(Hospital_Time.aov$residuals)
```




```
qqnorm(Hospital_Time.aov$fitted.values)
```



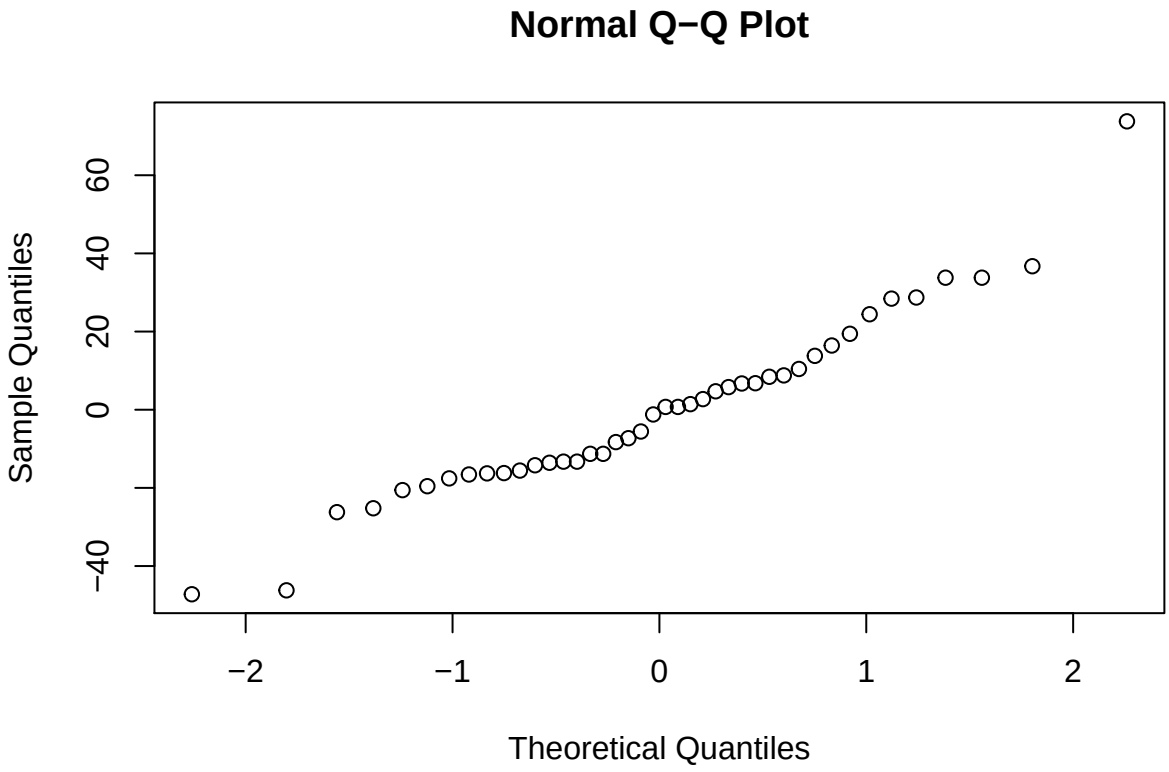

```
plot(Hospital_Time.aov$fitted.values, Hospital_Time.aov$residuals)
```




```
resid(Hospital_Time.aov)
```

```
##           1           2           3           4           5           6
##   5.7857143 -16.2142857 -46.2142857  -1.2142857   8.7857143  13.7857143
##           7           8           9          10          11          12
## -25.2142857  33.7857143  33.7857143  73.7857143 -26.2142857 -47.2142857
##          13          14          15          16          17          18
##   6.7857143 -14.2142857  28.4285714   8.4285714 -16.5714286   1.4285714
##          19          20          21          22          23          24
## -20.5714286  -5.5714286 -13.5714286 -19.5714286  24.4285714  16.4285714
##          25          26          27          28          29          30
##  19.4285714 -17.5714286  10.4285714 -15.5714286   4.7142857  36.7142857
##          31          32          33          34          35          36
## -16.2857143 -11.2857143   2.7142857   6.7142857   0.7142857 -13.2857143
##          37          38          39          40          41          42
##  28.7142857  -7.2857143   0.7142857 -13.2857143 -11.2857143  -8.2857143
```

```
qqnorm(resid(Hospital_Time.aov))
```




```
#Since it is linear, we can see the residuals are normally distributed, and thus it is a valid model.

bartlett.test(Waiting.time ~ Hospital, data = Hospital_Time)
```

```
##
##  Bartlett test of homogeneity of variances
##
## data:  Waiting.time by Hospital
## Bartlett's K-squared = 8.5166, df = 2, p-value = 0.01415
```

```
#Since P-Value is small, we cannot conlude that there are same variances.
#We can further prove this as max standard deviation is Hospital A with 33, and
#minimum standard deviation is Hospital C with 16. Therefore the ratio is greater than 2
#so it is not feasible to say it is the same variances.


kruskal.test(Waiting.time ~ Hospital, Hospital_Time)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  Waiting.time by Hospital
## Kruskal-Wallis chi-squared = 7.6051, df = 2, p-value = 0.02231
```

```
#Therefore, we can decisively conclude to reject the null hypothesis, as P-Value <
#0.05, and that one of the means is different to the others. Thus, we reject the null
#hypothesis under the kruskal test and the ANOVA test.
```