Homework (Class 12 Questions)

Destiny (A16340362)

Q13: Read this file into R and determine the sample size for each genotype and their corresponding median expression levels for each of these genotypes.

Sample size for genotype is 462 and the median expression is 25.11561.

Hint: The read.table(), summary() and boxplot() functions will likely be useful here. There is an example R script online to be used ONLY if you are struggling in vein. Note that you can find the medium value from saving the output of the boxplot() function to an R object and examining this object.

There is also the medium() and summary() function that you can use to check your understanding.

```
expression <- read.table("ExpressionGenotypeResult.txt")
head(expression)
```

```
  sample geno      exp
1 HG00367  A/G 28.96038
2 NA20768  A/G 20.24449
3 HG00361  A/A 31.32628
4 HG00135  A/A 34.11169
5 NA18870  G/G 18.25141
6 NA11993  A/A 32.89721
```
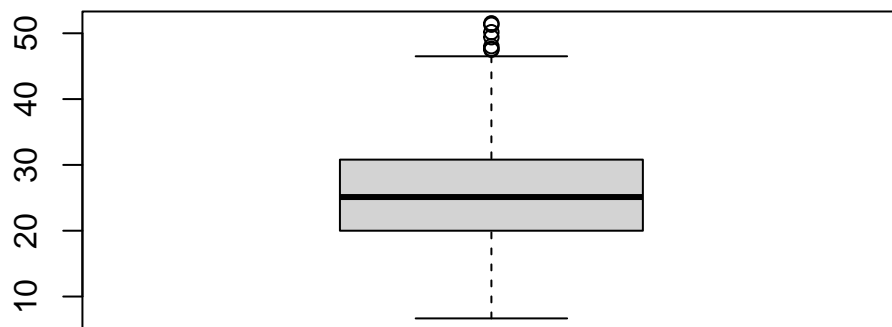
```
summary(expression)
```

```
   sample              geno               exp
 Length:462         Length:462         Min.   : 6.675
 Class :character   Class :character   1st Qu.:20.004
 Mode  :character   Mode  :character   Median :25.116
                                       Mean   :25.640
                                       3rd Qu.:30.779
                                       Max.   :51.518
```

```
table(expression$geno)
```

```
A/A A/G G/G
108 233 121
```

```
box <- boxplot(expression$exp)
```



```
median(box$stats[3])
```

```
[1] 25.11561
```

> Q14: Generate a boxplot with a box per genotype, what could you infer from the relative expression value between A/A and G/G displayed in this plot? Does the SNP effect the expression of ORMDL3?

Hint: An example boxplot is provided overleaf – yours does not need to be as polished as this one.

The SNP does effpect the expression of ORMDL3, since you can see with G/G there is a more reduced expression compared to having A/A.

```
library(ggplot2)
ggplot(expression) +aes(x=geno, y= exp)+ geom_boxplot()
```