class12

Q5

```
mxl <- read.csv("mxl.csv")
table(mxl$Genotype..forward.strand.)
```

```
A|A A|G G|A G|G
 22  21  12   9
```

```
nrow(mxl)
```

```
[1] 64
```

```
9/64
```

```
[1] 0.140625
```

Q13

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

```
# load up ggplot
library(ggplot2)
# read in data from txt file
table <- read.table("rs8067378_ENSG00000172057.6.txt")
# sample sizes per genotype
```

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

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

```
# make a boxplot, but no need to display it since we only need the values
box <- boxplot(table$exp ~ table$geno, plot=FALSE)
# retrieve stats of each box in the plot, specify row 3 for the medians
box$stats[3,]
```

```
[1] 31.24847 25.06486 20.07363
```

A/A: 108 samples, median 31.24847

A/G: 233 samples, median 25.06486

G/G: 121 samples, median 20.07363

Q14

Q: 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?

```
# create plot with axes for genotype (x) vs. expression (y)
p <- ggplot(table, aes(x=geno, y=exp)) +
        # make boxplot with coloring based on genotype so each box is a diff color
        geom_boxplot(aes(fill = geno), notch=TRUE) +
        # change axes titles/labels
        labs(x = "Genotype", y = "Expression")
p
```
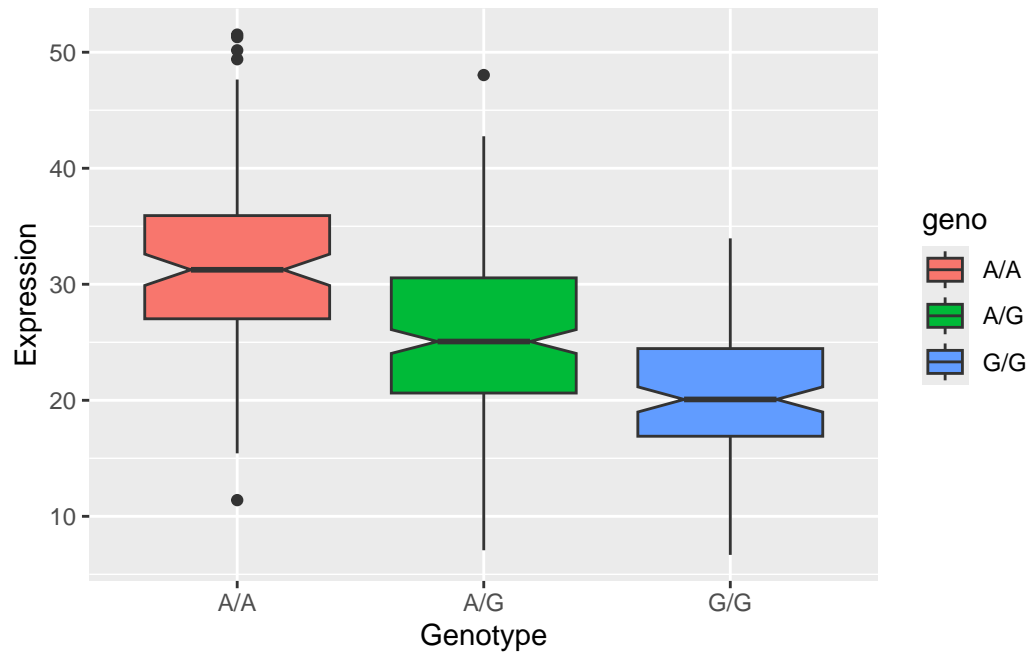
Looks like relative expression of A/A is generally greater (although a lot more spread out) compared to that of G/G in the plot. SNP does seem to affect the expression of ORMDL3.