Class 17 Questions 13&14

Dan Vu (PID: A17380158)

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.

```
expr <- read.table("rs8067378_ENSG00000172057.6.txt")
head(expr)
```

```
   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
```

```
nrow(expr)
```

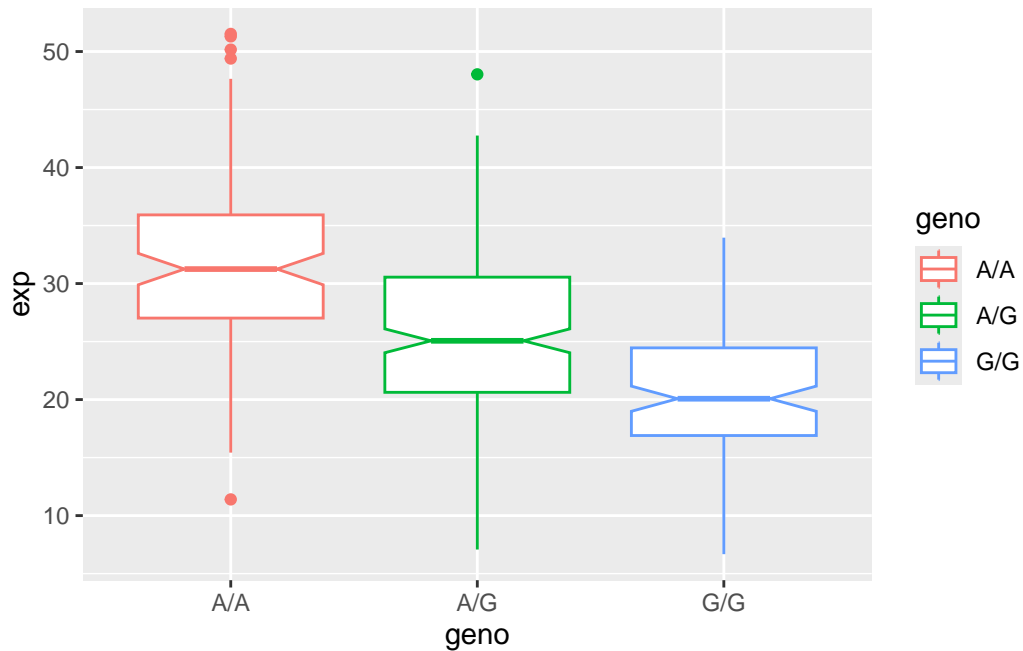```
[1] 462
```

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

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

I can infer from the plot created that A/A has a higher relative expression value compared to G/G. This means that the SNP does affect the expression of ORMDL3. There are 108 samples of A/A, 233 samples of A/G, and 121 samples of G/G.

```
library(ggplot2)
```

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?

```
ggplot(expr) +
  aes(geno, exp, col = geno) +
  geom_boxplot(notch = TRUE)
```


I can infer from the plot created that A/A has a higher relative expression value compared to G/G. This means that the SNP does affect the expression of ORMDL3.