

class 12 h2

Question 13

How many samples do we have?

```
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)
```

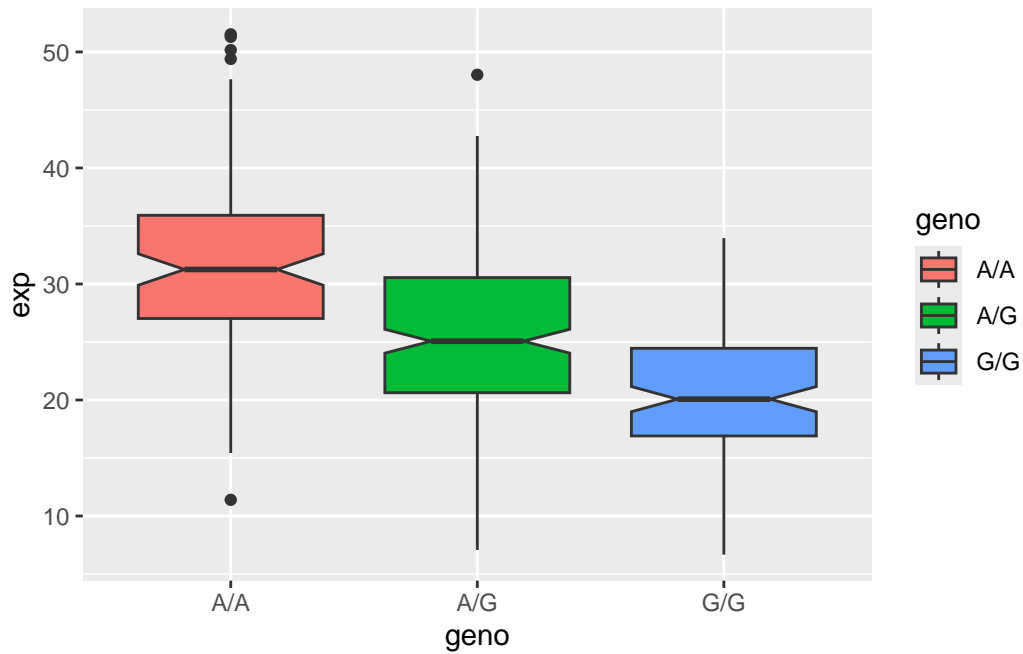
```
A/A A/G G/G
108 233 121
```

```
library(ggplot2)
```

Question 14

Let's make a boxplot

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



A/A has the highest median expression but has more variability while G/G has the lowest median expression but the least variability. The SNP does influence the expression of ORMDL3 because the A/A has higher expression while G/G has lower expression, which suggests that the G allele is associated with having a lower gene expression.