# Class 12

Josie (A11433761)

## Population Scale Analysis

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.

```r
expr<-read.table("sample_gen0_exp.csv")
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
```
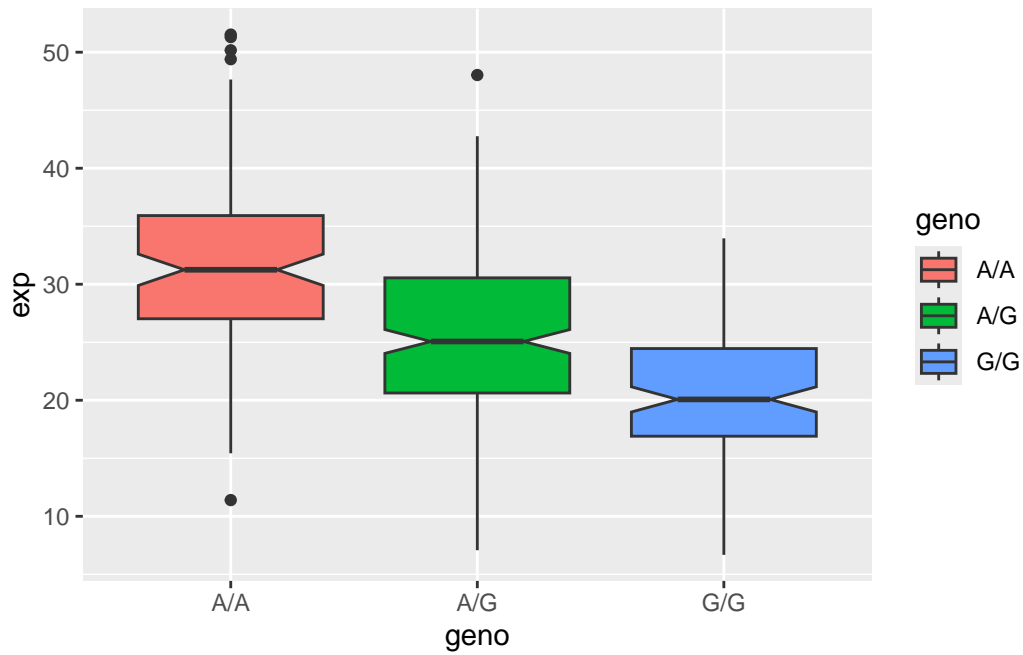
```r
nrow(expr)
```

```
[1] 462
```

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

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

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?

```
library(ggplot2)
ggplot(expr)+ aes(geno, exp, fill=geno)+
  geom_boxplot(notch=TRUE)
```



G/G in this location is associated the lower expression levels of ORMDL3.