

class-12-hw

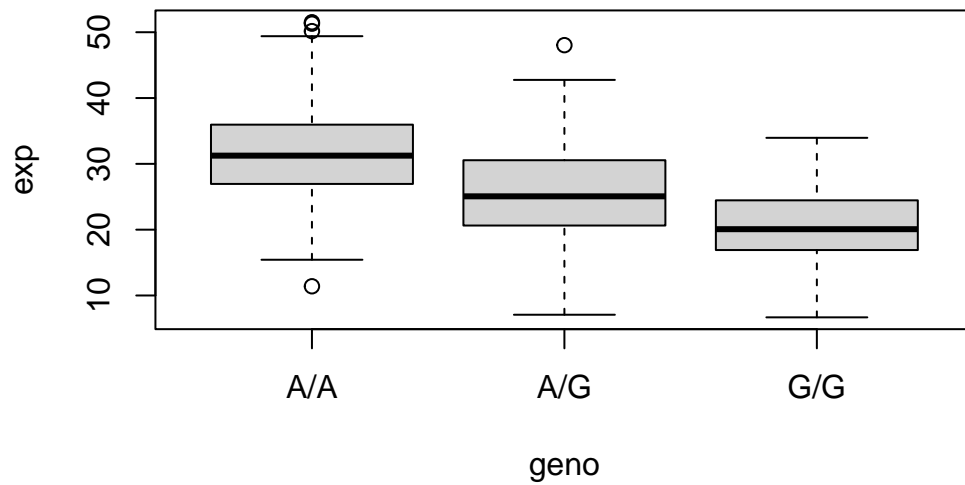
13.

```
data <- read.csv("/Users/jackie-l/Desktop/Bioinformatics/Class-11/class12results - Sheet1.csv")  
  
# i want genotype to be a factor  
data$geno <- as.factor(data$geno)  
summary(data)
```

sample	geno	exp
Length:462	A/A:108	Min. : 6.675
Class :character	A/G:233	1st Qu.:20.004
Mode :character	G/G:121	Median :25.116
		Mean :25.640
		3rd Qu.:30.779
		Max. :51.518

you can see that there is now a breakdown of sample size for each genotype

```
attempt <- boxplot(exp ~ geno, data = data)
```

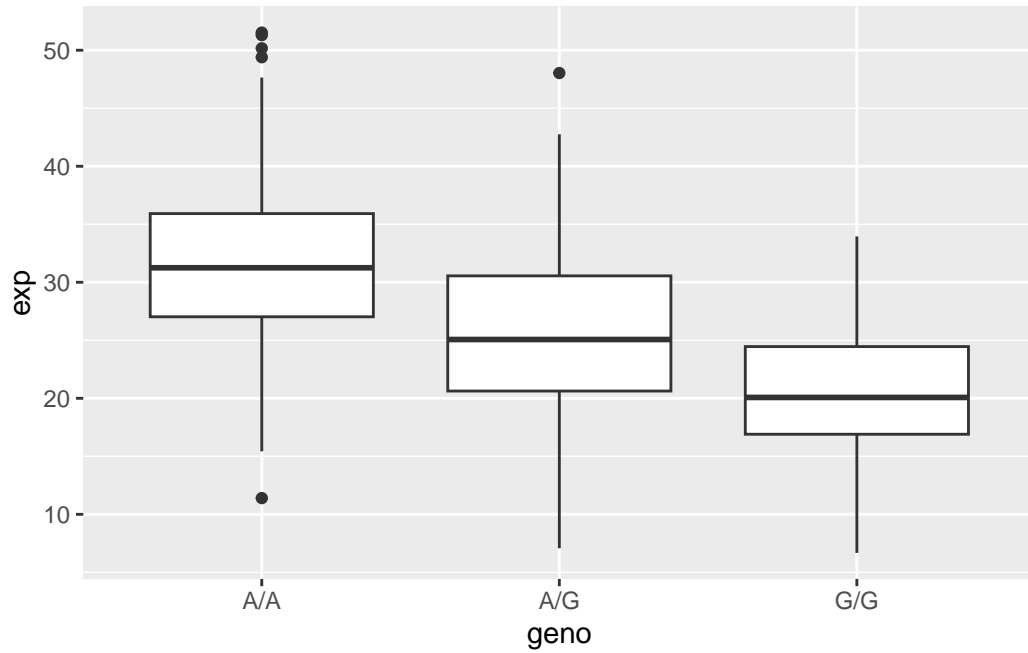


```
attempt$stats
```

```
      [,1]      [,2]      [,3]
[1,] 15.42908  7.07505  6.67482
[2,] 26.95022 20.62572 16.90256
[3,] 31.24847 25.06486 20.07363
[4,] 35.95503 30.55183 24.45672
[5,] 49.39612 42.75662 33.95602
```

Row 3 contains the median for the 3 groups- A/A 31.24847 A/G 25.06486 G/G 20.07363 14.

```
library(ggplot2)
plot<-ggplot(data,aes(geno,exp))+
  geom_boxplot()
plot
```



A/A is associated with higher ORMDL3 expression and GG has lower ORMDL3 expression.
So yes the SNP affects expression