

# Lab 12 genomics

AUTHOR

Tiffany Chin 15700705

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.

```
library(ggplot2)
file <- "https://bioboot.github.io/bggn213_W19/class-material/rs8067378_ENSG00000172057.6"
data <- read.table(file)
head(data)
```

	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

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

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

```
median(data$exp[data$geno=='A/A'])
```

```
[1] 31.24847
```

```
median(data$exp[data$geno=='A/G'])
```

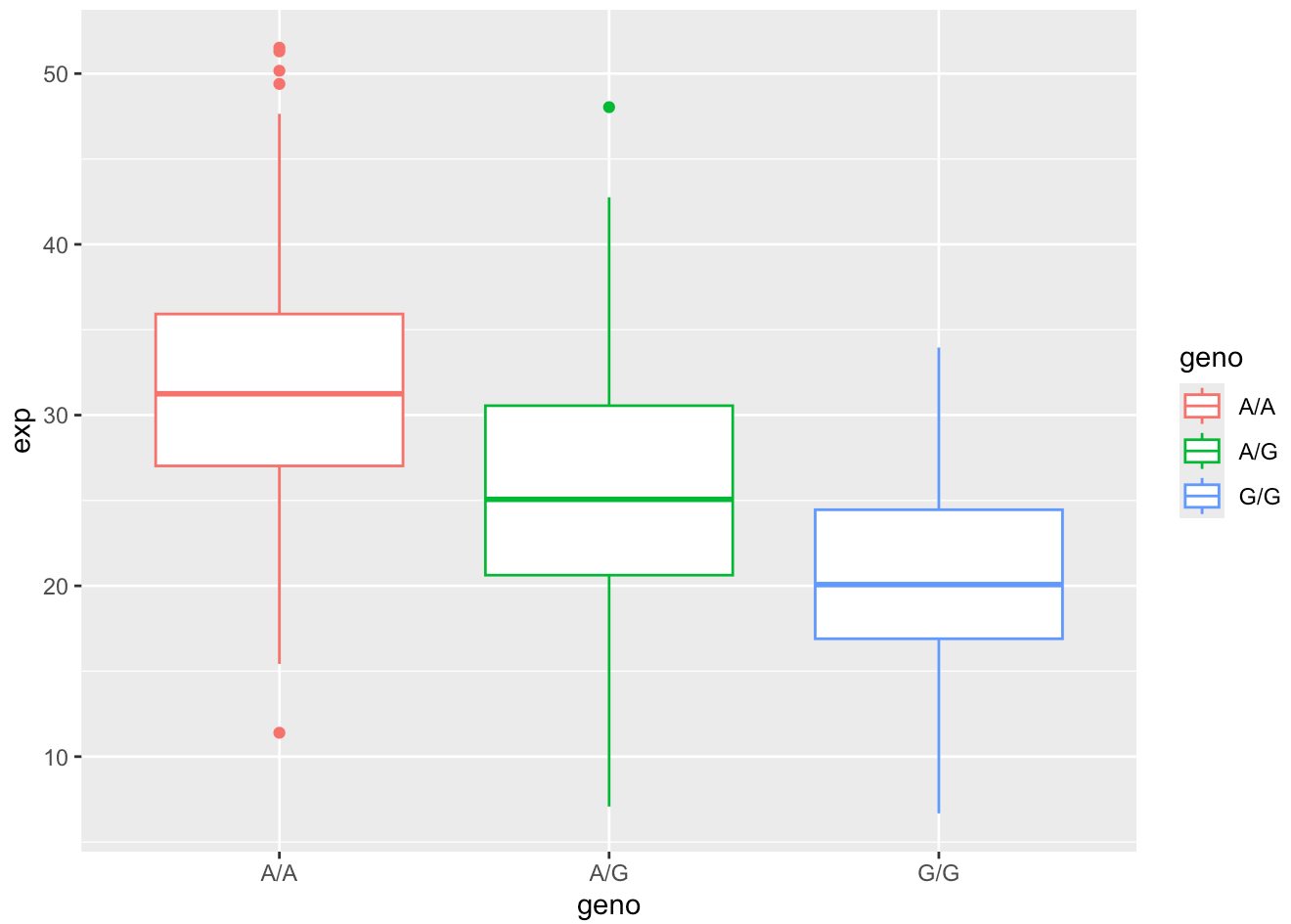
```
[1] 25.06486
```

```
median(data$exp[data$geno=='G/G'])
```

```
[1] 20.07363
```

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(data, aes(geno, exp, col = geno)) +
  geom_boxplot( )
```



Having a G/G SNP in this location is associated with a reduced expression of ORMDL3. Having an A/A or A/G in this location will result in higher expression of ORMDL3.