

class11

Zavier Annis

```
data <- data.frame(read.table("rs8067378_ENSG00000172057.6.txt"))
```

```
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

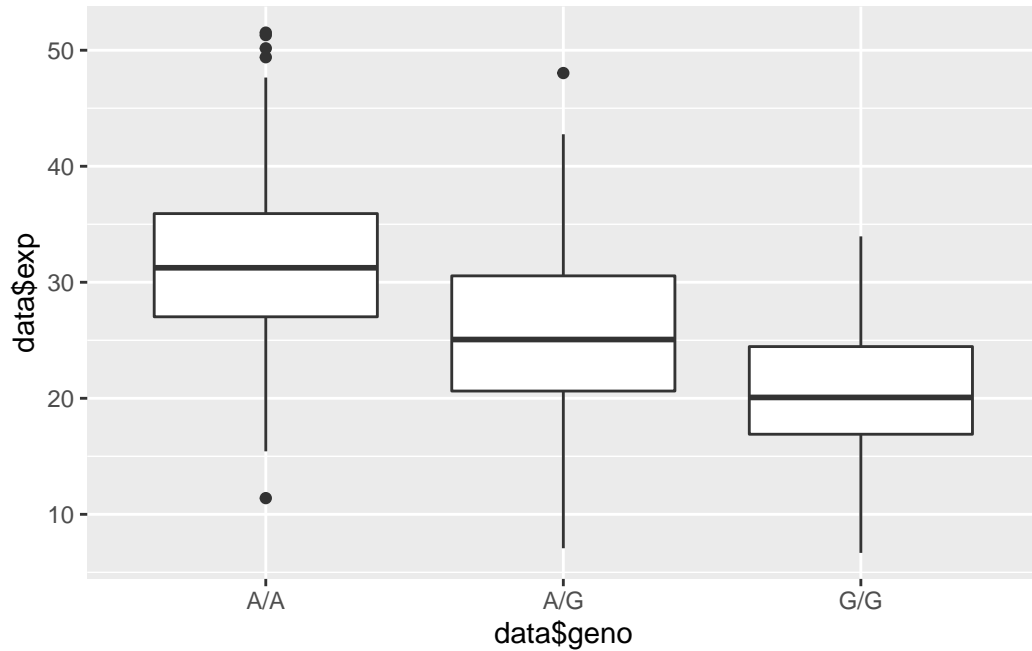
```
table(data$geno)
```

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

```
library(ggplot2)
boxplot <- ggplot(data, aes(x=data$geno, y=data$exp)) + geom_boxplot()
boxplot
```

Warning: Use of `data\$geno` is discouraged. Use `geno` instead.

Warning: Use of `data\$exp` is discouraged. Use `exp` instead.



A13. The sample sizes for each genotype are 108, 233, and 121 for A/A, A/G, and G/G, respectively. The median expression levels are ~31-32, 25, and 20, respectively.

A14. Yes, the SNP does appear to affect ORMDL3. The A/A genotype shows consistently higher expression of the gene than the G/G genotype, with a median more than 10 units higher and a lower quartile above the G/G upper quartile.