class11

Xueran Zou

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
csv <- read.table("rs8067378_ENSG00000172057.6.csv", header = TRUE, row.names = 1)
```

```r
table(csv$geno)
```

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

```r
library(dplyr)
```

```
Attaching package: 'dplyr'

The following objects are masked from 'package:stats':

    filter, lag

The following objects are masked from 'package:base':

    intersect, setdiff, setequal, union
```

```r
exp1 <- csv %>%
  filter(geno == "A/A")
median(exp1$exp)
```

```
[1] 31.24847
```

```
exp2 <- csv %>%
    filter(geno == "A/G")
median(exp2$exp)
```
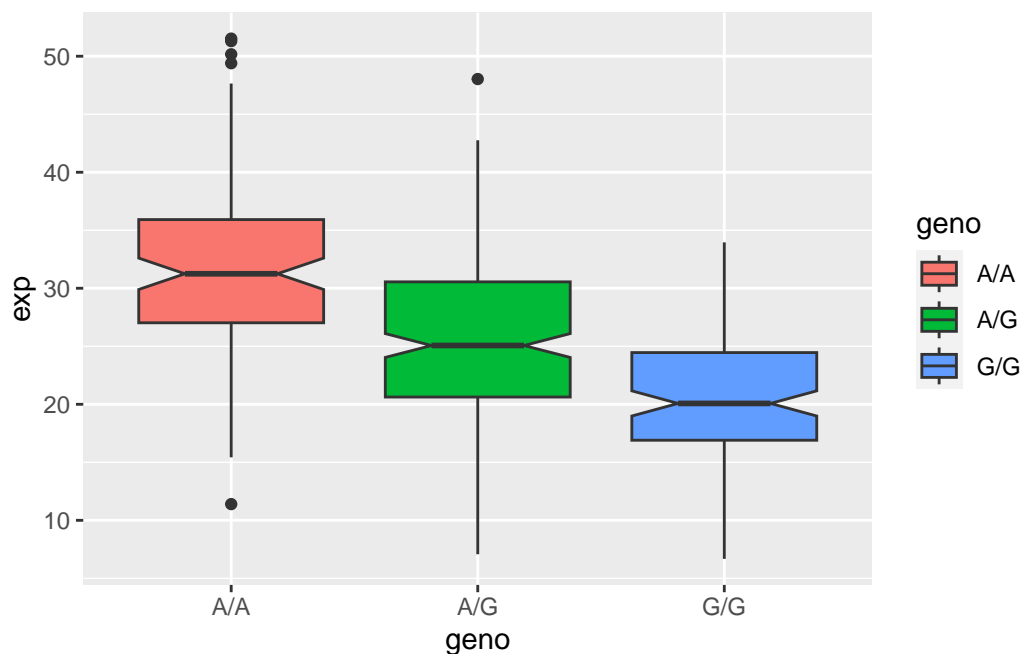
[1] 25.06486

```
exp3 <- csv %>%
    filter(geno == "G/G")
median(exp3$exp)
```

[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?

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


Having G/G is associated with the reduced expression of ORMDL3 compared to A/A.

The SNP effects the expression of ORMDL3.