Population Scale Analysis

Dora Deng (A17445600)

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.

read the data:

```
expr <- read.table("https://bioboot.github.io/bimm143_W25/class-material/rs8067378_ENSG00000
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
```

```
nrow(expr)
```

```
[1] 462
```
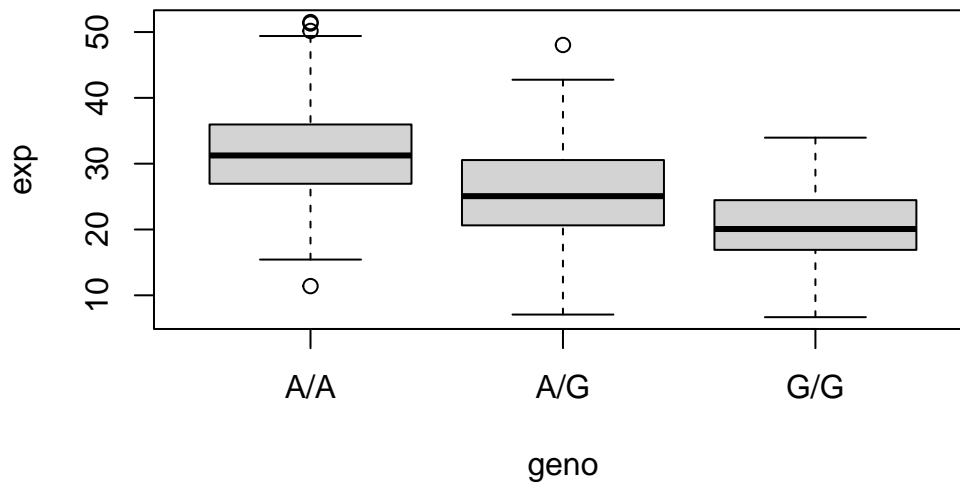
Figure out the sample size of each genotype.

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

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

```
library(ggplot2)
```

Use boxplot() to extract median expression levels

```
boxplot <- boxplot(exp ~ geno, expr)
```



```
boxplot_medians <- boxplot$stats[3, ]
names(boxplot_medians) <- levels(as.factor(expr$geno))
boxplot_medians
```

```
      A/A       A/G       G/G
31.24847 25.06486 20.07363
```

```
summary(expr$exp)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  6.675  20.004  25.116  25.640  30.779  51.518
```

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? Hint: An example boxplot is provided overleaf – yours does not need to be as polished as this one.

Lets make a boxplot.

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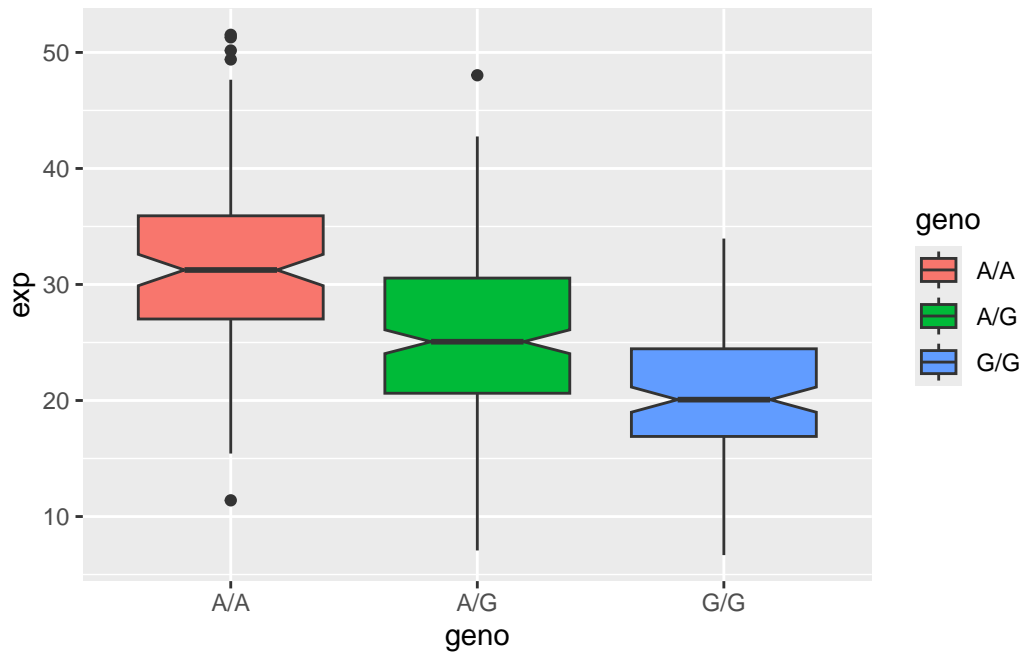

Since there are difference in the medium for each genotype, genotypes do affect the gene expression. SNP does affect the expression of ORMDL3, with A/A has a higher medium value and G/G has a comparatively lower medium value.