

# Class11HW

Mahsa Naeimi

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.

```
# First we need to read the data from url:
url <- "https://bioboot.github.io/bimm143_S18/class-material/rs8067378_ENSG00000172057.6.t
genot <- read.table(url)
```

```
# to find the sample size we use the summary function to get additional stats for each col
summary(genot)
```

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

There are 462 sample.

```
table(genot$geno)['G/G']
```

G/G  
121

From the data table we can see there are 3 genotype: G/G , A/A , G/A

```
# Sample size and Median expression level for G/G genotype
table(genot$geno)['G/G']
```

G/G  
121

```
summary( genot$exp[ genot$geno == "G/G" ] )
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
6.675	16.903	20.074	20.594	24.457	33.956

```
# Sample size and Median expression level for A/A genotype  
table(genot$geno)['A/A']
```

A/A  
108

```
summary( genot$exp[ genot$geno == "A/A" ] )
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
11.40	27.02	31.25	31.82	35.92	51.52

```
# Sample size and Median expression level for A/G genotype  
table(genot$geno)['A/G']
```

A/G  
233

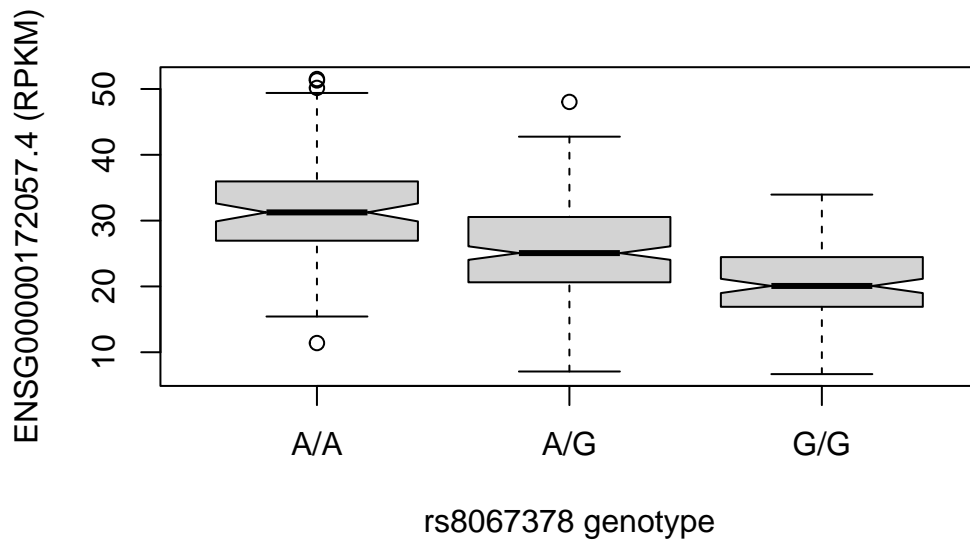
```
summary( genot$exp[ genot$geno == "A/G" ] )
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
7.075	20.626	25.065	25.397	30.552	48.034

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?

```
# Boxplot for each genotype expression data
```

```
p <- boxplot(exp~geno, data=genot, xlab="rs8067378 genotype", ylab="ENSG00000172057.4 (RPKM)
```



```
p
```

```
$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
```

```
$n
```

```
[1] 108 233 121
```

```
$conf
```

```
      [,1]      [,2]      [,3]
[1,] 29.87942 24.03742 18.98858
[2,] 32.61753 26.09230 21.15868
```

```

$out
[1] 51.51787 50.16704 51.30170 11.39643 48.03410

$group
[1] 1 1 1 1 2

$names
[1] "A/A" "A/G" "G/G"

```

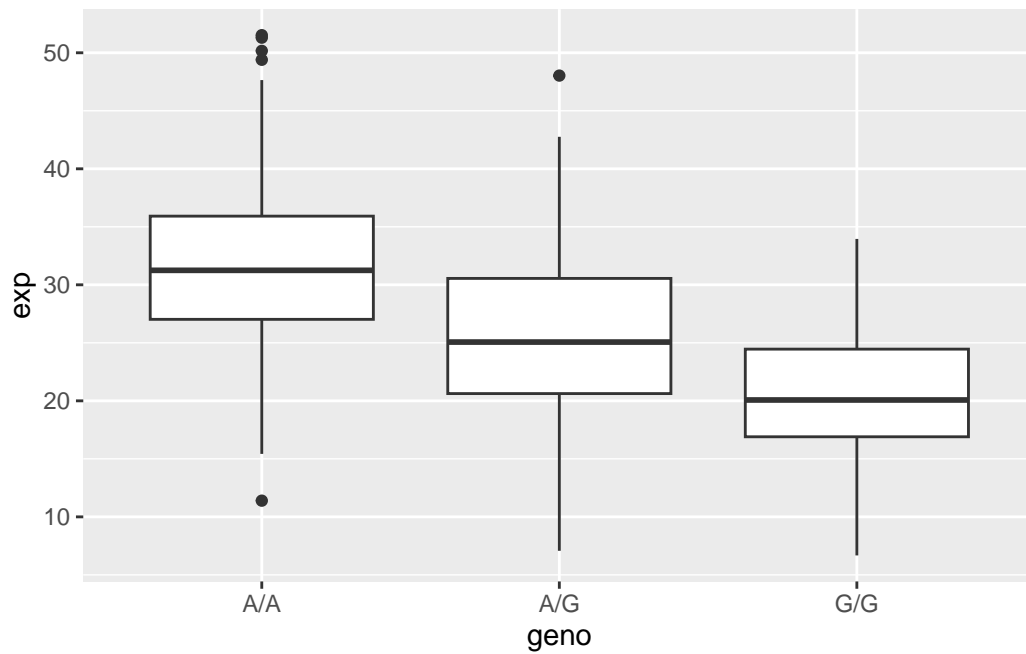
The boxplot shows that relative expression value of A/A is more than G/G. For this sample, this means higher expression level of ORMDL3 can happen when the genotype is A/A. So, SNP impacts the expression level of ORMDL3. someone with A/A genotype have higher expression level of ORMDL3 than someone with G/G genotype.

```

# to get a better visualization of expression level in each genotype we can use ggplot

library(ggplot2)
ggplot(genot, aes(genot, exp)) + geom_boxplot()

```



```
ggplot(genot, aes(geno, exp, fill=geno)) +
  geom_boxplot(notch=TRUE, outlier.shape = NA) +
  geom_jitter(shape=16, position=position_jitter(0.2), alpha=0.4)
```

