

Class 18 HW

Claire Chapman

12/2/2021

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.

```
url <- "https://bioboot.github.io/bggn213_F21/class-material/rs8067378_ENSG00000172057.6.txt"
data <- read.table(url)
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
```

```
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
```

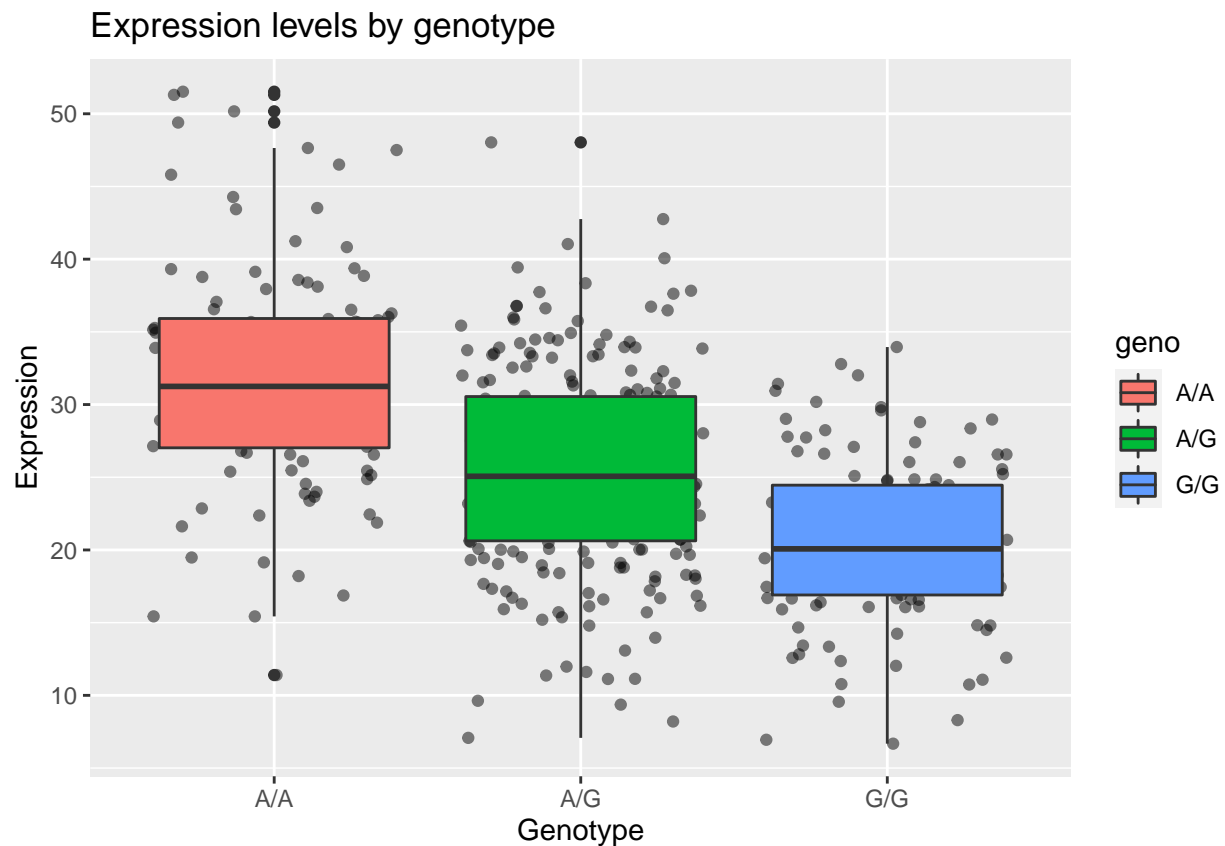
```
library(ggplot2)
```

```
data %>%
  group_by(genotype) %>%
  summarise(sample_size = n(), median = median(exp))
```

```
## # A tibble: 3 x 3
##   genotype sample_size median
##   <chr>      <int>    <dbl>
## 1 A/A         108     31.2
## 2 A/G         233     25.1
## 3 G/G         121     20.1
```

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?

```
data %>%
  group_by(geno) %>%
  ggplot(aes(geno, exp, fill = geno)) +
  geom_jitter(alpha = 0.5)+
  geom_boxplot()+
  labs(title = "Expression levels by genotype", x = "Genotype", y = "Expression")
```



Since there is a difference in expression between the A/A and G/G genotype, we can hypothesize that the SNP does affect the expression of ORMDL3.