

# Class 12

AUTHOR  
Erin Li

```
file_path = '/mnt/data/rs8067378_ENSG00000172057.6.csv'

# read the file
data <- read.table("rs8067378_ENSG00000172057.6.csv", header = TRUE, sep = "")

# Display the first few rows of the dataframe to understand its structure
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

```
summary_data <- data %>%
  group_by(geno) %>%
  summarise(SampleSize = n(),
            MedianExpression = median(exp, na.rm = TRUE))
```

Q13

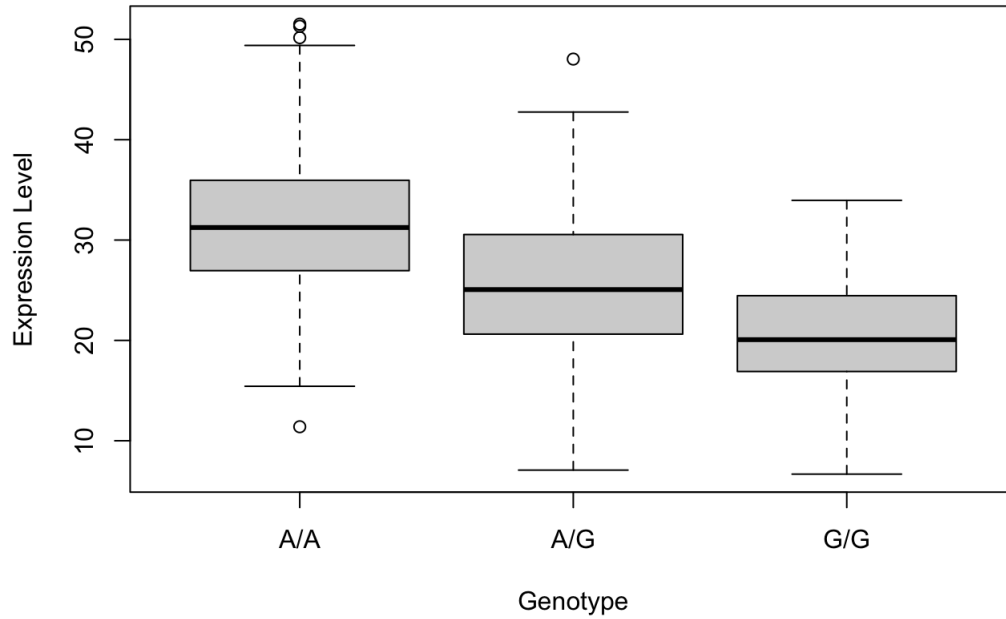
```
# Print the summary data
print(summary_data)
```

```
# A tibble: 3 × 3
  geno SampleSize MedianExpression
<chr>    <int>         <dbl>
1 A/A      108          31.2
2 A/G      233          25.1
3 G/G      121          20.1
```

Q14

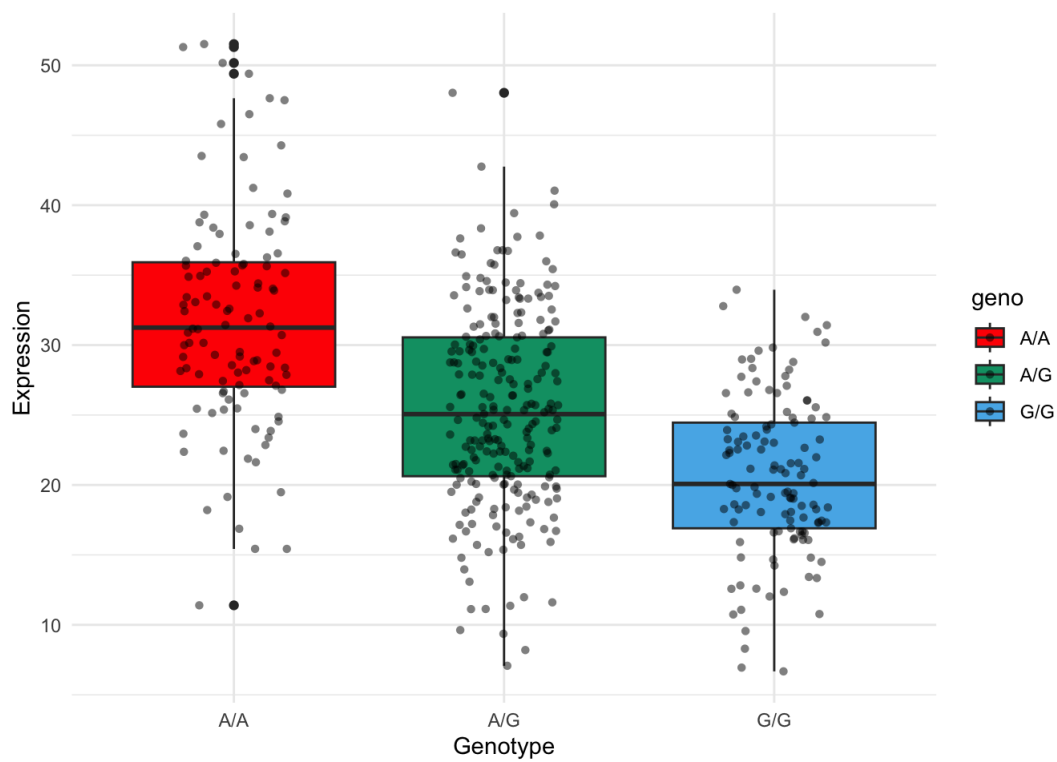
```
boxplot(exp ~ geno, data = data, main = "Expression Levels by Genotype",
        xlab = "Genotype", ylab = "Expression Level")
```

## Expression Levels by Genotype



```
library(ggplot2)

ggplot(data, aes(x=geno, y=exp, fill=geno)) +
  geom_boxplot() +
  geom_jitter(shape=16, position=position_jitter(0.2), alpha=0.5) + # Adds the points on the plot
  scale_fill_manual(values=c("#FF0000", "#009E73", "#56B4E9")) + # Change colors as needed
  labs(x="Genotype", y="Expression") +
  theme_minimal()
```



The median expression level for the genotype A/A is higher than for G/G. The interquartile range (IQR), which represents the middle 50% of the data, is narrower for G/G than for A/A, suggesting that expression levels for G/G are more consistent compared to A/A. There are some outliers in both groups, but especially in the A/A group. These patterns suggest the A/A genotype may be linked to increased ORMDL3 expression compared to G/G.