

# Homework Class 11

Julia Ainsworth

## Table of contents

<b>Section 4: Population Scale Analysis [HOMEWORK]</b>	<b>1</b>
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. . . . .	1
<b>Q14: Generate a boxplot with a box per genotype, what could you infer from the relative</b>	<b>2</b>

## Section 4: Population Scale Analysis [HOMEWORK]

**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.**

```
data <- read.table("https://bioboot.github.io/bggn213_W19/class-material/rs8067378_ENSG000  
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
```

```
nrow(data)
```

```
[1] 462
```

Sample size:

```
table(data$geno)
```

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

```
summary(data$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? It seems like A/A has higher expression than G/G

```
library(ggplot2)
```

Making a boxplot:

```
ggplot(data) + aes(geno, exp, fill =geno) + geom_boxplot(notch = T)
```

