Class12

Jae Kim

Section 1. Proportion of G/G in a population

```
mxl <- read.csv("MXL.csv", row.names = 1)
mxl
```

```
           Genotype..forward.strand. Population.s. Father Mother
NA19648 (F)                      A|A ALL, AMR, MXL      -      -
NA19649 (M)                      G|G ALL, AMR, MXL      -      -
NA19651 (F)                      A|A ALL, AMR, MXL      -      -
NA19652 (M)                      G|G ALL, AMR, MXL      -      -
NA19654 (F)                      G|G ALL, AMR, MXL      -      -
NA19655 (M)                      A|G ALL, AMR, MXL      -      -
NA19657 (F)                      A|G ALL, AMR, MXL      -      -
NA19658 (M)                      A|A ALL, AMR, MXL      -      -
NA19661 (M)                      A|G ALL, AMR, MXL      -      -
NA19663 (F)                      A|A ALL, AMR, MXL      -      -
NA19664 (M)                      G|A ALL, AMR, MXL      -      -
NA19669 (F)                      A|A ALL, AMR, MXL      -      -
NA19670 (M)                      A|A ALL, AMR, MXL      -      -
NA19676 (M)                      G|G ALL, AMR, MXL      -      -
NA19678 (F)                      A|A ALL, AMR, MXL      -      -
NA19679 (M)                      A|G ALL, AMR, MXL      -      -
NA19681 (F)                      A|G ALL, AMR, MXL      -      -
NA19682 (M)                      A|G ALL, AMR, MXL      -      -
NA19684 (F)                      A|G ALL, AMR, MXL      -      -
NA19716 (F)                      G|A ALL, AMR, MXL      -      -
NA19717 (M)                      A|G ALL, AMR, MXL      -      -
NA19719 (F)                      G|G ALL, AMR, MXL      -      -
NA19720 (M)                      G|G ALL, AMR, MXL      -      -
NA19722 (F)                      G|A ALL, AMR, MXL      -      -
```

```
NA19723 (M)                    G|G ALL, AMR, MXL        -        -
NA19725 (F)                    A|G ALL, AMR, MXL        -        -
NA19726 (M)                    A|A ALL, AMR, MXL        -        -
NA19728 (F)                    A|A ALL, AMR, MXL        -        -
NA19729 (M)                    A|G ALL, AMR, MXL        -        -
NA19731 (F)                    A|A ALL, AMR, MXL        -        -
NA19732 (M)                    A|G ALL, AMR, MXL        -        -
NA19734 (F)                    G|A ALL, AMR, MXL        -        -
NA19735 (M)                    G|G ALL, AMR, MXL        -        -
NA19740 (F)                    A|A ALL, AMR, MXL        -        -
NA19741 (M)                    A|A ALL, AMR, MXL        -        -
NA19746 (F)                    A|A ALL, AMR, MXL        -        -
NA19747 (M)                    G|A ALL, AMR, MXL        -        -
NA19749 (F)                    A|G ALL, AMR, MXL        -        -
NA19750 (M)                    A|G ALL, AMR, MXL        -        -
NA19752 (F)                    A|G ALL, AMR, MXL        -        -
NA19755 (F)                    A|A ALL, AMR, MXL        -        -
NA19756 (M)                    G|A ALL, AMR, MXL        -        -
NA19758 (F)                    A|G ALL, AMR, MXL        -        -
NA19759 (M)                    G|A ALL, AMR, MXL        -        -
NA19761 (F)                    G|A ALL, AMR, MXL        -        -
NA19762 (M)                    A|A ALL, AMR, MXL        -        -
NA19764 (F)                    A|A ALL, AMR, MXL        -        -
NA19770 (F)                    A|G ALL, AMR, MXL        -        -
NA19771 (M)                    A|A ALL, AMR, MXL        -        -
NA19773 (F)                    A|A ALL, AMR, MXL        -        -
NA19774 (M)                    A|G ALL, AMR, MXL        -        -
NA19776 (F)                    A|G ALL, AMR, MXL        -        -
NA19777 (M)                    A|A ALL, AMR, MXL        -        -
NA19779 (F)                    G|A ALL, AMR, MXL        -        -
NA19780 (M)                    A|A ALL, AMR, MXL        -        -
NA19782 (F)                    G|A ALL, AMR, MXL        -        -
NA19783 (M)                    A|G ALL, AMR, MXL        -        -
NA19785 (F)                    A|A ALL, AMR, MXL        -        -
NA19786 (M)                    G|A ALL, AMR, MXL        -        -
NA19788 (F)                    A|G ALL, AMR, MXL        -        -
NA19789 (M)                    G|G ALL, AMR, MXL        -        -
NA19792 (M)                    A|A ALL, AMR, MXL        -        -
NA19794 (F)                    G|A ALL, AMR, MXL        -        -
NA19795 (M)                    A|G ALL, AMR, MXL        -        -
```

```r
table(mxl$Genotype..forward.strand.)
```

```
A|A A|G G|A G|G
 22  21  12   9
```

```r
table(mxl$Genotype..forward.strand.) / nrow(mxl)*100
```

```
     A|A     A|G     G|A     G|G
 34.3750 32.8125 18.7500 14.0625
```

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

```r
url <- "https://bioboot.github.io/bimm143_S23/class-material/rs8067378_ENSG00000172057.6.t

expr <- read.table(url)

nrow(expr)
```

```
[1] 462
```

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

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

```r
result <- aggregate(expr$exp ~ expr$geno, data = expr, FUN = median)
result
```
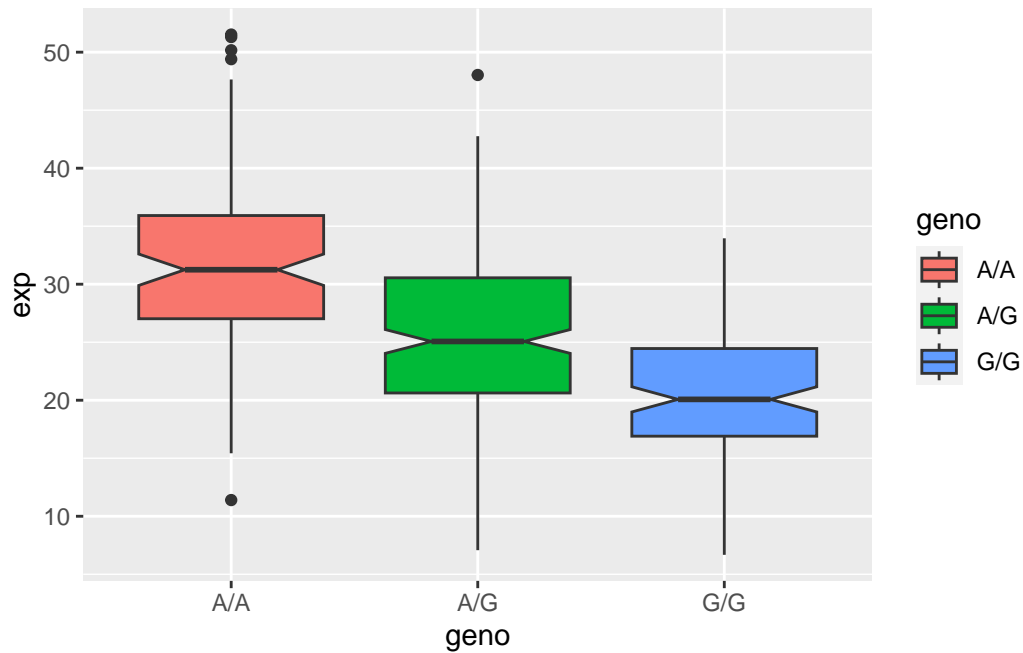
```
  expr$geno expr$exp
1       A/A 31.24847
2       A/G 25.06486
3       G/G 20.07363
```

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?

```r
library(ggplot2)
```

Making a boxplot

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



```r
boxplot_data <- ggplot_build(x)$data[[1]]
boxplot_data$middle
```

```
[1] 31.24847 25.06486 20.07363
```

From the boxplot, we can see that the expression of ORMDL3 is higher if the genotype is A/A than if it is G/G. It seems like SNP does affect the expression of ORMDL3.