

class11

```
library(readxl)
mxl_snp_var <- read.csv("373531-SampleGenotypes-Homo_sapiens_Variation_Sample_rs8067378.csv")

table(mxl_snp_var$Genotype..forward.strand.)
```

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

```
length(mxl_snp_var$Genotype..forward.strand.)
```

```
[1] 64
```

```
library(dplyr)
```

Warning: package 'dplyr' was built under R version 4.1.1

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

```
counts <- count(mx1_snp_var, Genotype..forward.strand.)
```

Q5. Proportion of the sample population with G|G:

```
counts[4,2]/sum(counts[,2])
```

```
[1] 0.140625
```

Homework from 2/15/23

```
snp_pop <- read.csv("snp_pop_exp.txt", sep = " ")
```

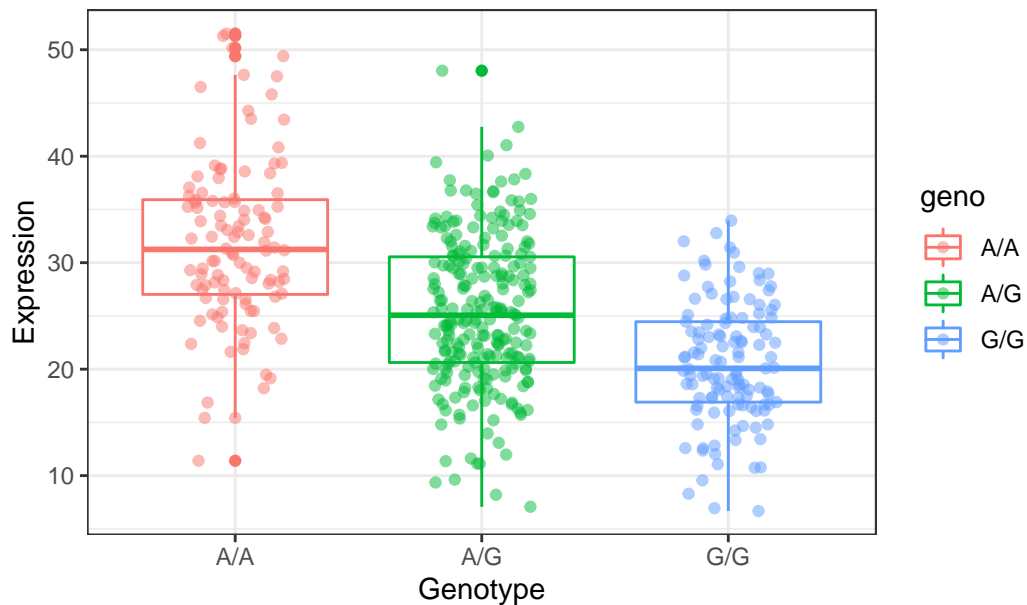
```
table(snp_pop$geno)
```

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

Plot the genotypes

```
library(ggplot2)
ggplot(data = snp_pop) +
  aes(x = geno, y = exp, group = geno, color = geno) +
  geom_boxplot(stat = "boxplot") +
  geom_jitter(alpha = 0.5, width = 0.2) +
  theme_bw() +
  labs(title = "Expression levels of ORM DL3 by genotype of rs8067378", x = "Genotype", y =
```

Expression levels of ORMDL3 by genotype of rs8067378



```
summary(snp_pop)
```

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

Getting the summary stats:

```
inds <- snp_pop$geno == "A/G"
```

Above is a logical vector containing the SNP of choice (A/G)

Stats for Expression for A/G SNP

```
summary(snp_pop$exp[inds=="TRUE"])
```

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

Repeat above for A/A G/A and G/G