class19

Thisha Thiagarajan A15474979

12/3/2021

Introduction to Genome Informatics Lab

Question 13/14

```r
expr <- read.table("rs8067378_ENSG00000172057.6.txt")
head(expr)
```

```
##     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
```

```r
nrow(expr)
```

```
## [1] 462
```
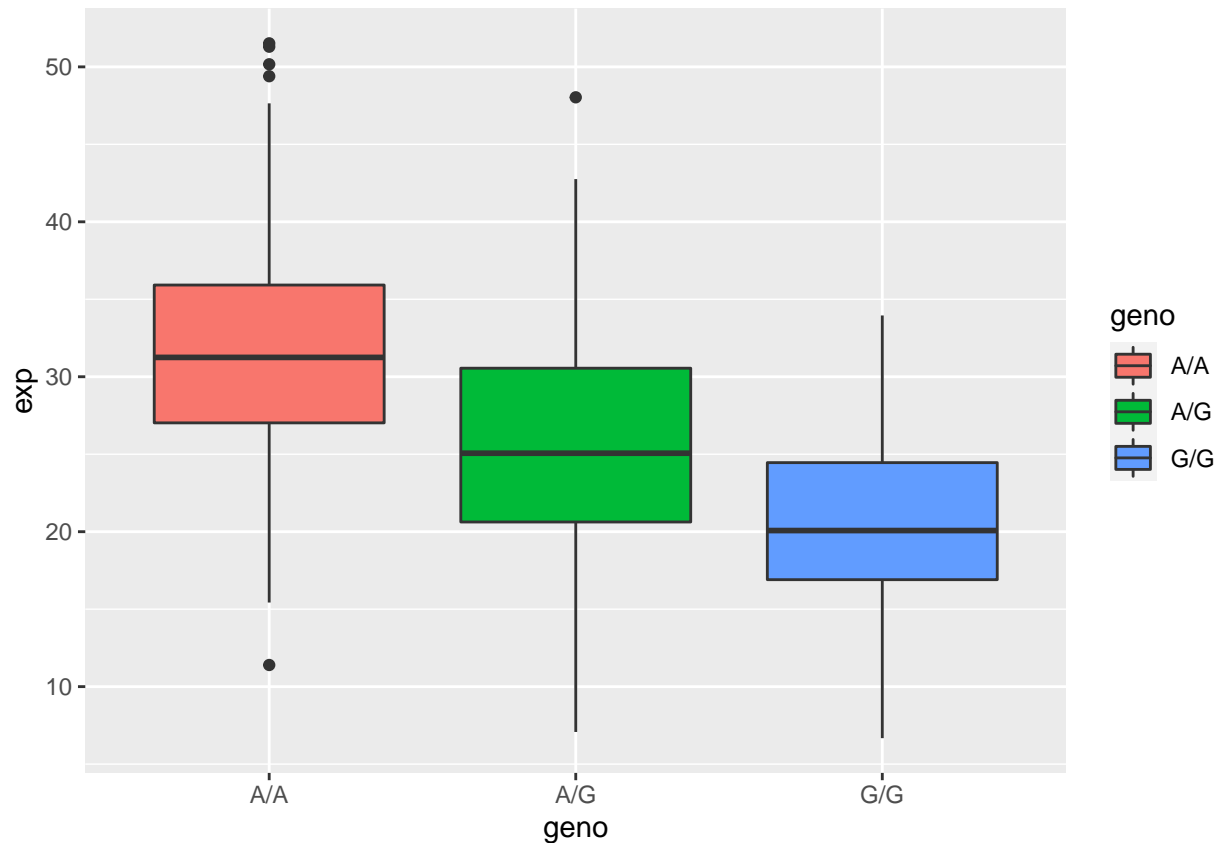
We have 462 samples. Found using nrow().

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

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

Using table(), I was able to determine sample size for each genotype. A/A = 108, A/G 233, G/G = 121.

Now, I am going to make a boxplot to visualize this data.

```r
library(ggplot2)
g <- ggplot(expr) + aes(x = geno, y = exp, fill = geno) +
  geom_boxplot()

g
```

Looking at this, you can tell that the median for A/A is around 32, the median for A/G is around 25, and the median for G/G is around 20.

```
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
```

```
AA <- filter(expr, geno == "A/A")
mean(AA$exp)
```

```
## [1] 31.81864
```

```
median(AA$exp)
```

```
## [1] 31.24847
```

```
AG <- filter(expr, geno == "A/G")
mean(AG$exp)
```

```
## [1] 25.3968
```

```
median(AG$exp)
```

```
## [1] 25.06486
```

```
GG <- filter(expr, geno == "G/G")
mean(GG$exp)
```

```
## [1] 20.59371
```

```
median(GG$exp)
```

```
## [1] 20.07363
```