

Project10

Haroon Riyaz (PID A15377799)

2/21/2022

Section 1. Proportion of G/G in population

Downloaded CSV file from Ensemble

Read CSV file

```
mx1 <- read.csv("373531-SampleGenotypes-Homo_sapiens_Variation_Sample_rs8067378.csv")
head(mx1)
```

```
##      Sample..Male.Female.Unknown. Genotype..forward.strand. Population.s. Father
## 1              NA19648 (F)              A|A ALL, AMR, MXL      -
## 2              NA19649 (M)              G|G ALL, AMR, MXL      -
## 3              NA19651 (F)              A|A ALL, AMR, MXL      -
## 4              NA19652 (M)              G|G ALL, AMR, MXL      -
## 5              NA19654 (F)              G|G ALL, AMR, MXL      -
## 6              NA19655 (M)              A|G ALL, AMR, MXL      -
##      Mother
## 1      -
## 2      -
## 3      -
## 4      -
## 5      -
## 6      -
```

```
table(mx1$Genotype..forward.strand.)
```

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

Calculate proportion

```
round(table(mx1$Genotype..forward.strand.)/nrow(mx1) * 100, 2)
```

```
##
## A|A A|G G|A G|G
## 34.38 32.81 18.75 14.06
```

Section 4: Population Scale Analysis

How many samples do we have?

```
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
```

```
# How many samples there are
nrow(expr)
```

```
## [1] 462
```

How many of each type?

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

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

Visualize data

```
library(ggplot2)
```

Make Boxplot

```
ggplot(expr) + aes(geno, exp, fill = geno) + geom_boxplot(notch = TRUE)
```

