

# BIMM143class12

## Section 1 Proportion of G/G in a population

#Downloaded a CSV file from Ensemble[https://www.ensembl.org/Homo\\_sapiens/Variation/Sample?db=core;r=17:39894595-39895595;v=rs8067378;vdb=variation;vf=105535077;sample=MXL#373531\\_tablePanel](https://www.ensembl.org/Homo_sapiens/Variation/Sample?db=core;r=17:39894595-39895595;v=rs8067378;vdb=variation;vf=105535077;sample=MXL#373531_tablePanel)

Here we read this CSV file

```
mxl<-read.csv("class12.csv")
head(mxl)
```

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

A A	A G	G A	G G
22	21	12	9

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

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

```
##Section 4 Population Scale Analysis
```

```
##How many samples do we have?
```

```
expr<-read.table("BIMM143class12optional.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
```

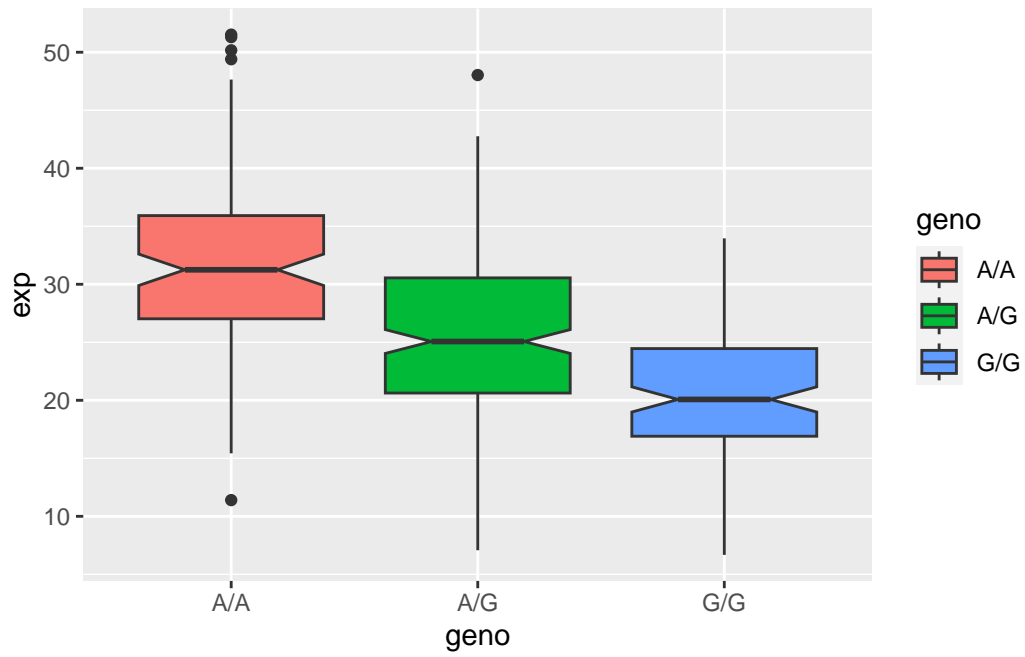
```
nrow(expr)
```

```
[1] 462
```

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

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

```
library(ggplot2)  
##Lets make a boxplot  
##you do not even need x=, y=  
ggplot(expr) +  
  aes(geno, exp, fill=geno) +  
  geom_boxplot(notch=TRUE)
```



*##notch adds "belt" in the midle*