

Lab 12: Genomics

Trevor Hoang (A16371830)

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

Downloaded a CSV from Ensemble and then using that CSV for the data

```
mxl = read.csv("373531-SampleGenotypes-Homo_sapiens_Variation_Sample_rs8067378.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(mx1$Genotype..forward.strand.)
```

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

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

	A A	A G	G A	G G
	0.343750	0.328125	0.187500	0.140625

##Section 4. Population Scale Analysis

One sample is not enough to know what is happening in the entire population. To we need to process more samples to see if there is any association of the 4 asthma associated SNPs on ORMDL3 expression.

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

```
nrow(expr)
```

```
[1] 462
```

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

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

```
library(ggplot2)

#Making a boxplot
ggplot(expr)+
  aes(geno,exp, fill=geno)+
  geom_boxplot(notch=T)
```

