

# Class 17: Intro to Genome Informatics

Shazreh Hassan (A13743949)

## Table of contents

Section 1 . . . . .	1
---------------------	---

### Section 1

Q13: Read this file into R and determine the sample size for each genotype and their corresponding median expression levels for each of these genotypes.

```
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)
```

Q14: Lets make a boxplot:

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

