

HWk7

A17576411

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.

##A: sample size: A|A = 108, A|G = 283, G|G = 121 medians: A|A = 31.2, A|G = 25.06, G|G = 20.07

```
data = read.table("rs8067378_ENSG00000172057.6.txt")
head(data)
```

	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

```
summary(data)
```

sample	geno	exp
Length:462	Length:462	Min. : 6.675
Class :character	Class :character	1st Qu.:20.004
Mode :character	Mode :character	Median :25.116
		Mean :25.640
		3rd Qu.:30.779
		Max. :51.518

```
sampsize <- table(data$geno)
print(sampsize)
```

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

```
median <- tapply(data$exp,data$geno,median)
print(median)
```

A/A	A/G	G/G
31.24847	25.06486	20.07363

Q14: Generate a boxplot with a box per genotype, what could you infer from the relative expression value between A/A and G/G displayed in this plot? Does the SNP effect the expression of ORM DL3?

A: From the boxplot, it appears that there is less expression of G/G relative to A/A. Since there is this drop in expression between A/A and G/G, it's obvious that any SNPS will decrease the expression of the gene.

```
bp <- boxplot(exp ~ geno, data=data, main = "Boxplot")
```

