

hw_genomics

Mari Williams (A15858833)

```
df <- read.table("https://bioboot.github.io/bggn213_W19/class-material/rs8067378_ENSG00000177170")
head(df)
```

	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

###Q13

```
summary(df)
```

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

```
table(df$geno)
```

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

108 A/A, 233 A/G, and 121 G/G.

```
library(dplyr)
```

Warning: package 'dplyr' was built under R version 4.3.3

Attaching package: 'dplyr'

The following objects are masked from 'package:stats':

filter, lag

The following objects are masked from 'package:base':

intersect, setdiff, setequal, union

```
df %>%  
  group_by(geno) %>%  
  summarise(mean_exp = mean(exp))
```

```
# A tibble: 3 x 2  
  geno mean_exp  
  <chr>   <dbl>  
1 A/A      31.8  
2 A/G      25.4  
3 G/G      20.6
```

###Q14

```
library(ggplot2)  
  
ggplot(df, aes(x = geno, y = exp)) +  
  geom_boxplot() +  
  geom_jitter(width = 0.1, alpha = 0.5) +  
  labs(x = "Genotype", y = "Expression", title = "Expression by genotype")
```

