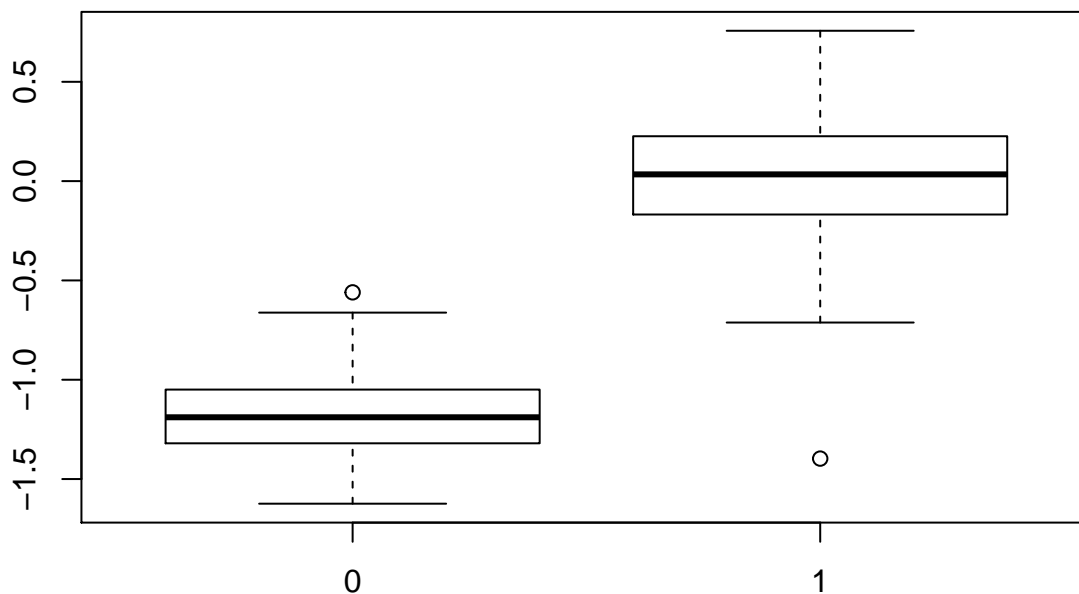# Exercise 7

Your Name

11 Dec 2015

```r
genes <- read.delim("gene.description.txt")
subjects <- read.delim("cancer.patients.txt")
evals <- read.delim("gene.expression.txt",stringsAsFactors = FALSE)
```

```r
ind <- match("ESR1", genes[,2])
probe <- genes[ind,1]
genevals <- evals[match(probe,rownames(evals)),]
```

```r
boxplot(as.numeric(genevals)~factor(subjects$er))
```



```r
t.test(as.numeric(genevals)~factor(subjects$er))
```

```
##
##  Welch Two Sample t-test
##
## data:  as.numeric(genevals) by factor(subjects$er)
## t = -38.746, df = 205.88, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -1.246953 -1.126198
## sample estimates:
## mean in group 0 mean in group 1
##     -1.17388506      0.01269076
```

- write the expression matrix for just ER negative samples to a file

```
erNegPatients <- subjects$er == 0
write.table(evals[,erNegPatients], file="erNegativeExpression.txt")
```

- create a new data frame combining the patient information with ESR1 expression level

```
df <- cbind(subjects, ER = as.numeric(genevals))
head(df)
```

```
##          samplename age er grade     ER
## NKI_4         NKI_4  41  1     3 -0.007
## NKI_6         NKI_6  49  1     2  0.074
## NKI_7         NKI_7  46  0     1 -0.767
## NKI_8         NKI_8  48  0     3 -0.820
## NKI_9         NKI_9  48  1     3 -0.180
## NKI_11       NKI_11  37  1     3 -0.296
```