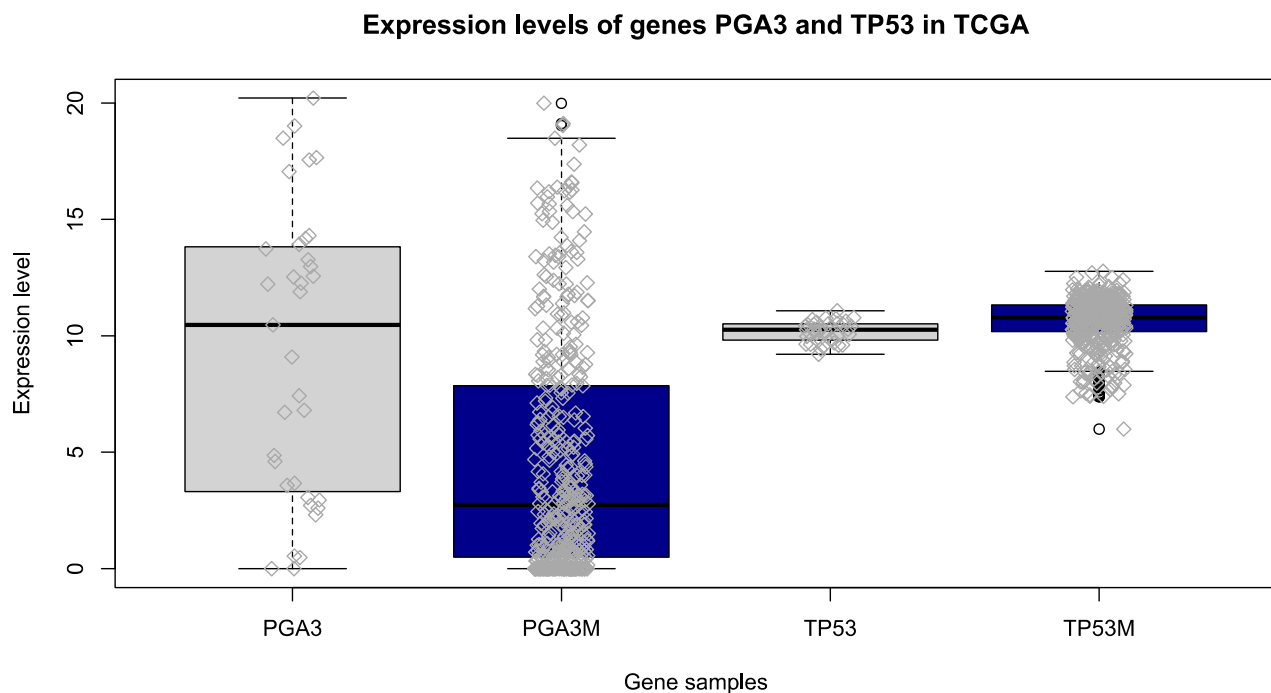


## Midterm

The following boxplots describe the expression in normal gastric samples and tumor gastric samples. The boxplots present the expression levels of gene PGA3 and TP53 in TCGA samples. We can see that there are four plots in the graph, which are PGA3 normal sample and the tumor sample, as well as TP53 normal sample and the tumor sample. We can see that in the first box plot which is the normal PGA3 sample, the expression level of the gene is quite high comparing to the second box plot, the tumor example. From observation, we see that the normal PGA3 sample is more active than the tumor sample, so we can conclude that less active PGA3 could be related the functions of the digestive system. The next two boxplots are the normal sample and the tumor sample from gene TP53, and we can see that the gene is more active in the tumor sample. TP53 gene provides instructions for making a protein called tumor protein p53, which helps to prevent the development of tumors and regulate cell division and that explains why the expression of TP53 is higher in the tumor sample.



## R Code:

```
getwd()
setwd("C:/Users/sing_/Desktop/2016 Fall/ModernAppDS/mid term/")

mydata <- repeatable("stad.txt", row.names=1, header=TRUE, sep="\t",
strip.white=TRUE)
mydata2 <- log2(mydata +1)

mydata3 <- read.table("stad-t.txt", row.names=1, header=TRUE,
sep="\t", strip.white=TRUE)
mydata4 <- log2(mydata3 + 1)

boxplot(list(t(mydata2[1,]), t(mydata4[1,]), t(mydata2[2,]),
t(mydata4[2,])), names=c("PGA3", "PGA3M", "TP53", "TP53M"),
main="Expression levels of genes PGA3 and TP53 in TCGA", xlab="Normal
and mutation", ylab="mutation data", col=(c("purple", "gold"))))

stripchart(list(t(mydata2[1,]), t(mydata4[1,]), t(mydata2[2,]),
t(mydata4[2,])), vertical = TRUE, method = "jitter", pch = 5, col =
"grey", bg = "bisque", add = TRUE)
```