

# Homework 1

*Jugal Marfatia, Namrata Ray, Md Muhtasim Billah*

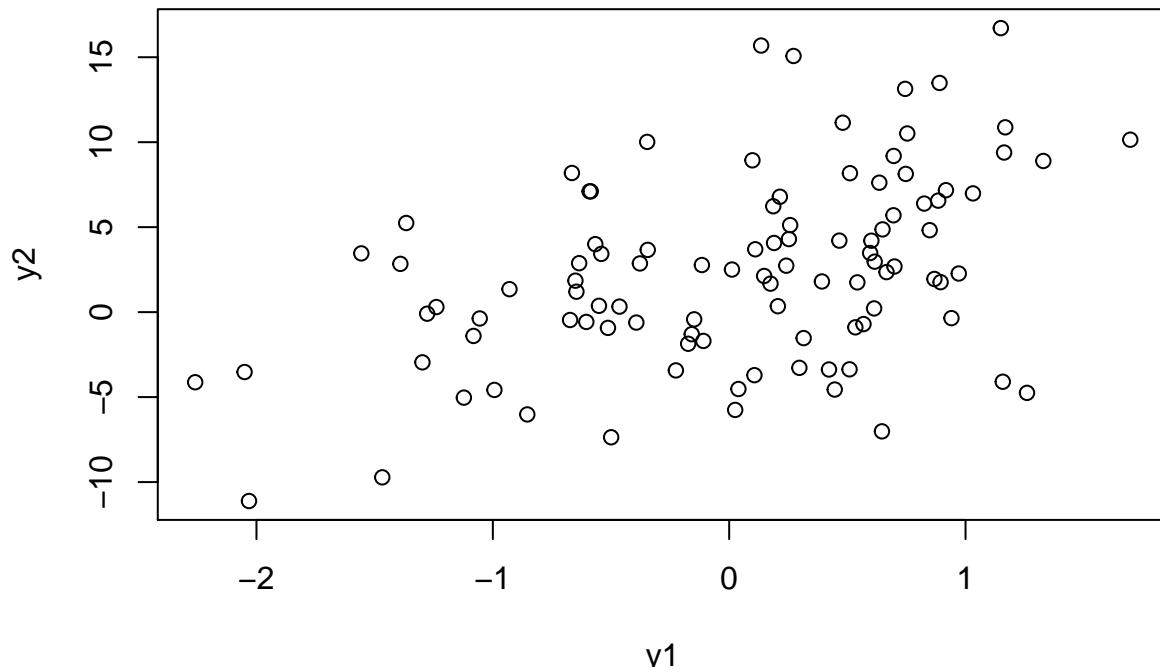
*10/01/2019*

## Question 1

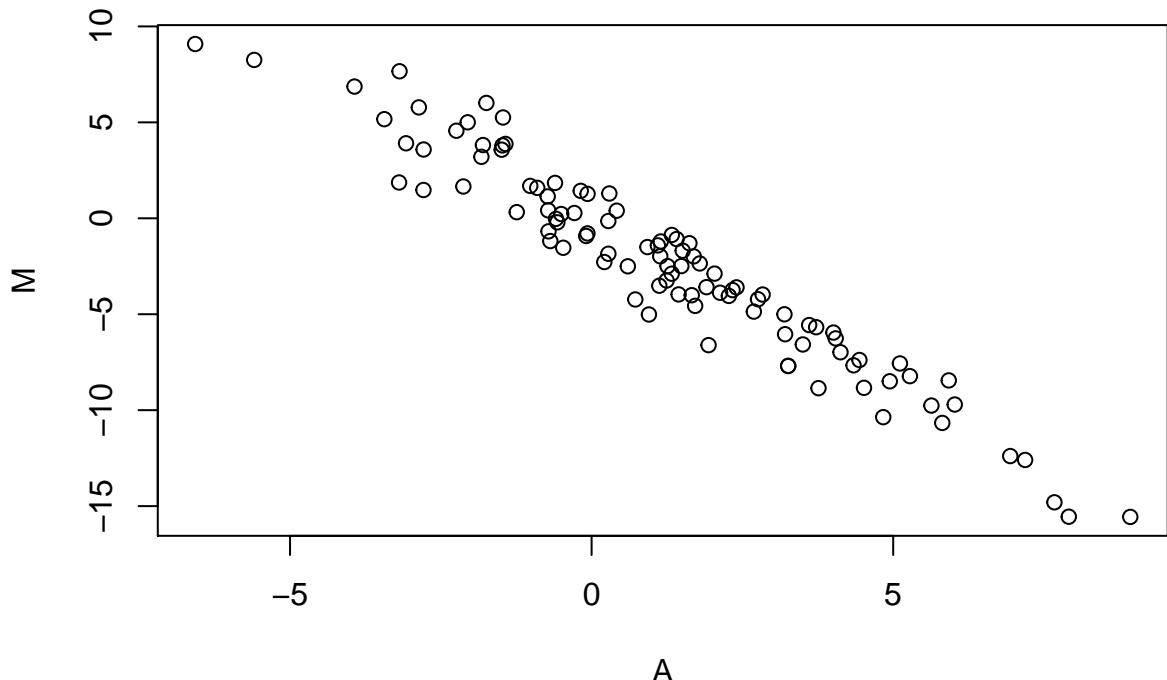
I have put the data set loess2.csv on the website. Use the data to run loess plots using spans of .1, .3, .5, .7 and .9. Give the loess plot with the different spans using the lines option in R.

```
data=read.csv("loessdata2.csv",header=TRUE)
#head(data)
y1=data$x2
y2=data$y2
M=y1-y2
A=(y1+y2)/2

plot(y1,y2)
```



```
plot(A,M)
```



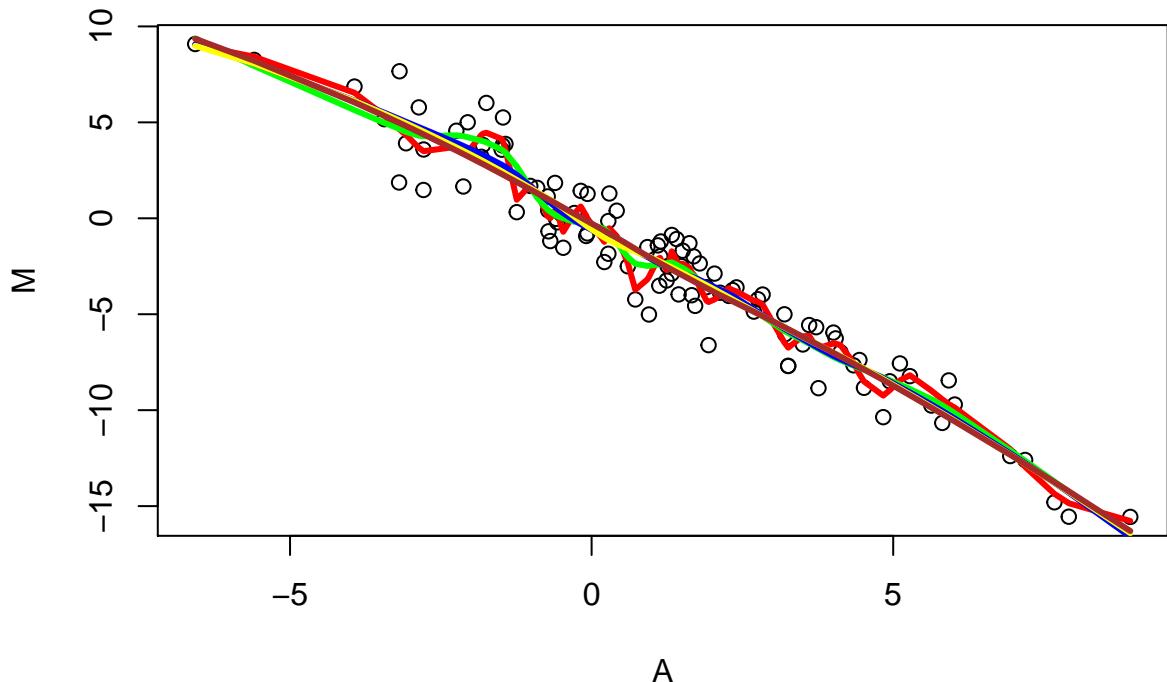
```
#loess plots with the different spans

j <- order(A)

lw1 <- loess(M ~ A,data=data,span=0.10)
lw2 <- loess(M ~ A,data=data,span=0.30)
lw3 <- loess(M ~ A,data=data,span=0.50)
lw4 <- loess(M ~ A,data=data,span=0.70)
lw5 <- loess(M ~ A,data=data,span=0.90)

#plotting all the lines together

plot(M ~ A, data=data)
lines(A[j],lw1$fitted[j],col="red",lwd=3)
lines(A[j],lw2$fitted[j],col="green",lwd=3)
lines(A[j],lw3$fitted[j],col="blue",lwd=3)
lines(A[j],lw4$fitted[j],col="yellow",lwd=3)
lines(A[j],lw5$fitted[j],col="brown",lwd=3)
```



## Question 2

Consider the following are data from 3 arrays 6 spots each. Give the quantile-normalized array values.

```

array1 = c(5.390,    12.074,     3.247,      4.274,      11.24,   14.861)
array2 = c(5.813,    8.777,     2.914,      2.736,      6.398,   23.020)
array3 = c(7.522,    7.719,     1.783,      10.59,     4.439,   21.972)

array1_s = sort(array1)
array2_s = sort(array2)
array3_s = sort(array3)

matrix = rbind(c(5.390, 12.074,     3.247,      4.274,      11.24,   14.861),
               c(5.813,  8.777,     2.914,      2.736,      6.398,   23.020),
               c(7.522,  7.719,     1.783,      10.59,     4.439,   21.972))

matrix_s = rbind(array1_s, array2_s, array3_s)
c_means = colMeans(matrix_s)

for(i in 0:length(array1_s)+1){
  matrix[matrix == array1_s[i]] = c_means[i]
  matrix[matrix == array2_s[i]] = c_means[i]
  matrix[matrix == array3_s[i]] = c_means[i]
}
matrix

##          [,1]      [,2]      [,3]      [,4]      [,5]      [,6]
## [1,] 6.241667 10.480333 2.588667  3.875667 8.452333 19.951
## [2,] 6.241667 10.480333 3.875667  2.588667 8.452333 19.951
## [3,] 6.241667  8.452333 2.588667 10.480333 3.875667 19.951

```

### Question 3

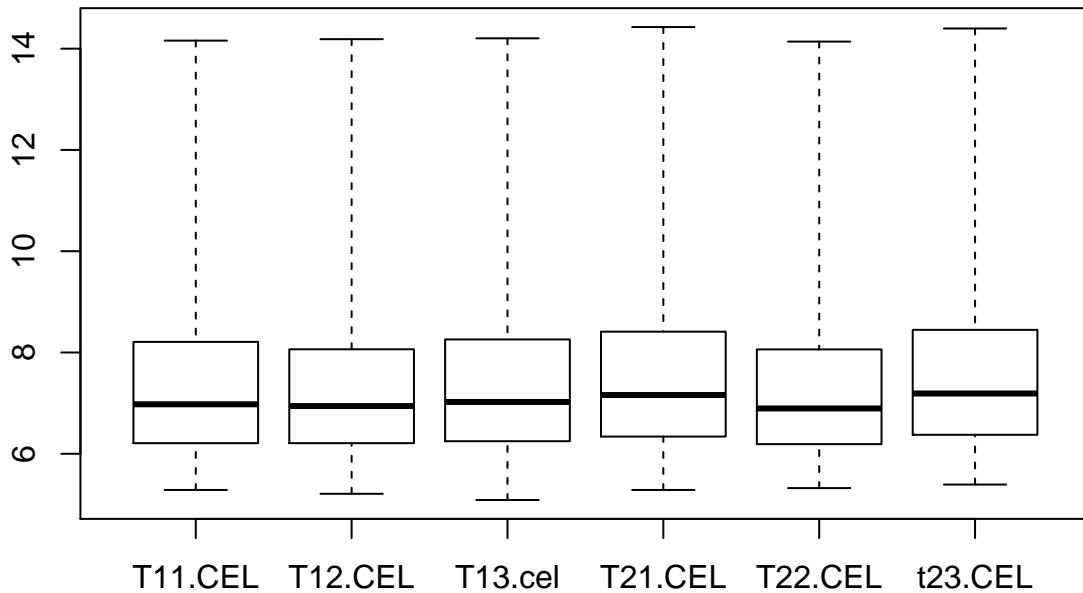
I have put the data set for the craig.zip on the website. Download the files and do M versus A plots, boxplots and histograms with and without normalization. Use the methods discussed in class, MAS5 and RMA for normalization.

```
library(affyPLM)
library(preprocessCore)
a.data=ReadAffy()
```

#### Boxplot without normalization

```
boxplot(a.data)
```

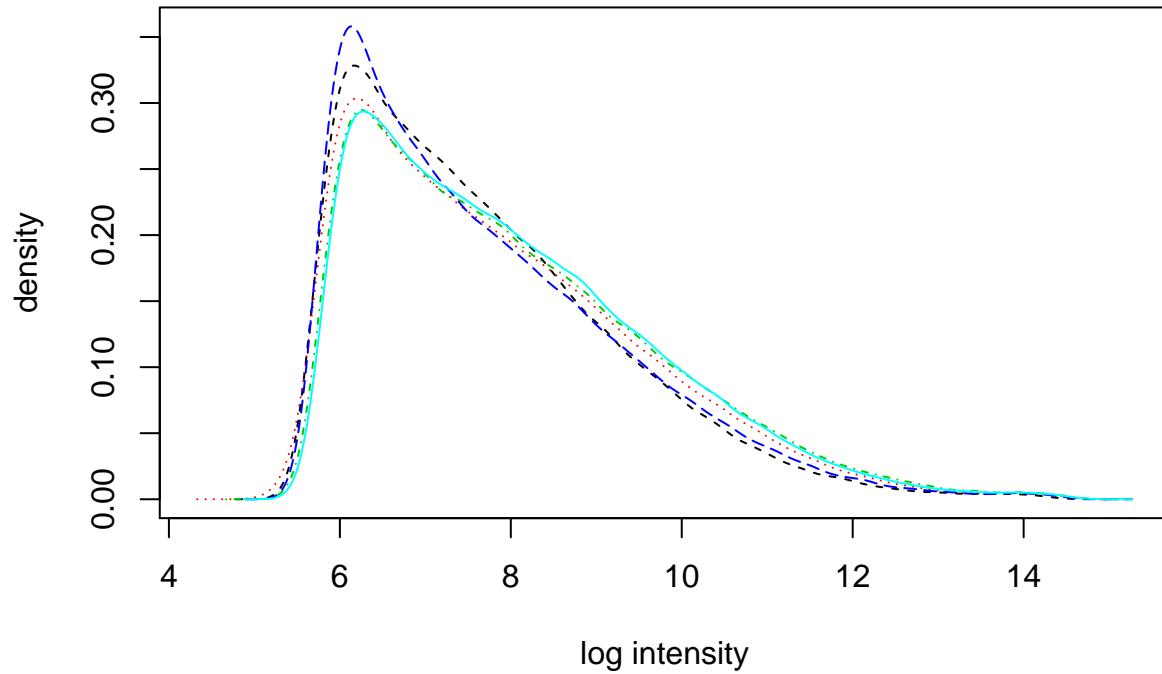
```
##
```



#### Histogram without normalization

```
hist(a.data,col=0:6,main="Histogram of probe intensities")
```

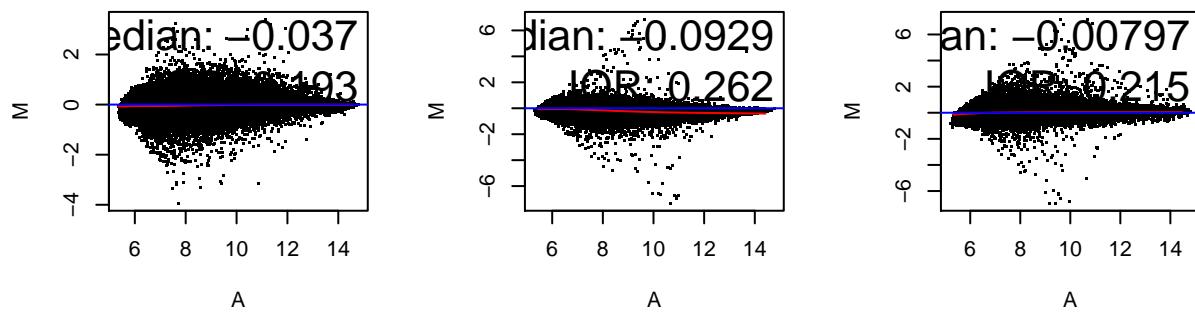
## Histogram of probe intensities



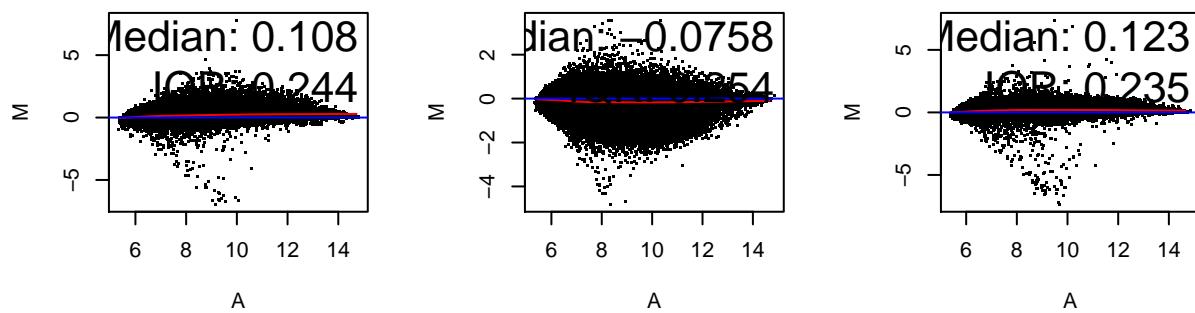
MAplot without normalization

```
par(mfrow=c(2,3))
MAplot(a.data)
```

1.CEL vs pseudo-median reference.CEL vs pseudo-median reference3.cel vs pseudo-median reference



1.CEL vs pseudo-median reference.CEL vs pseudo-median reference1.CEL vs pseudo-median reference

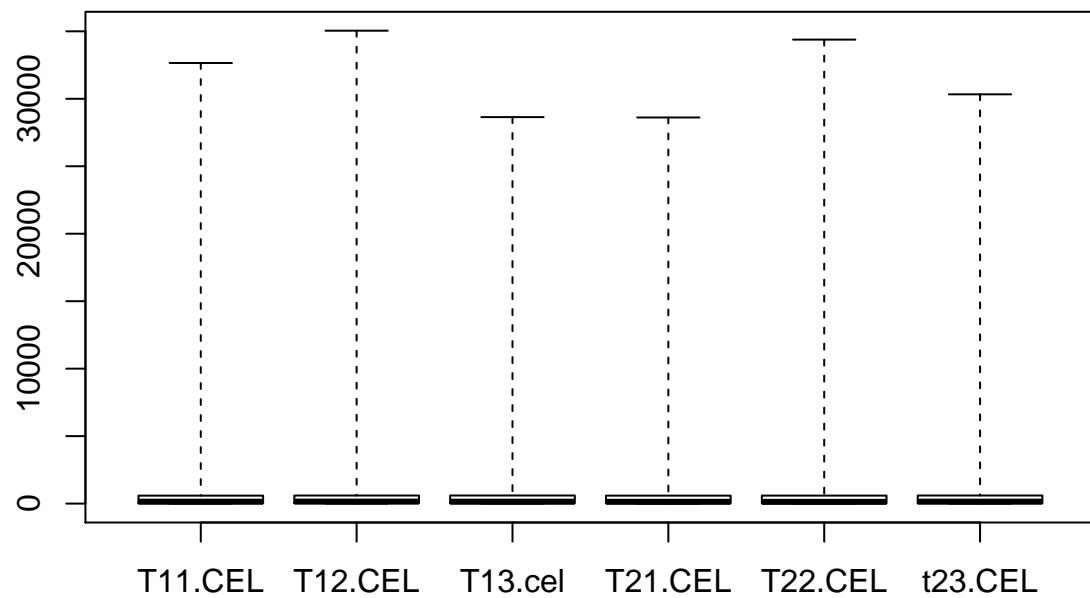


Boxplot with MAS5 normalization

```
eset.mas5=mas5(a.data)

## background correction: mas
## PM/MM correction : mas
## expression values: mas
## background correcting...done.
## 24128 ids to be processed
## |                               |
## |#####|#####|#####|#####|
```

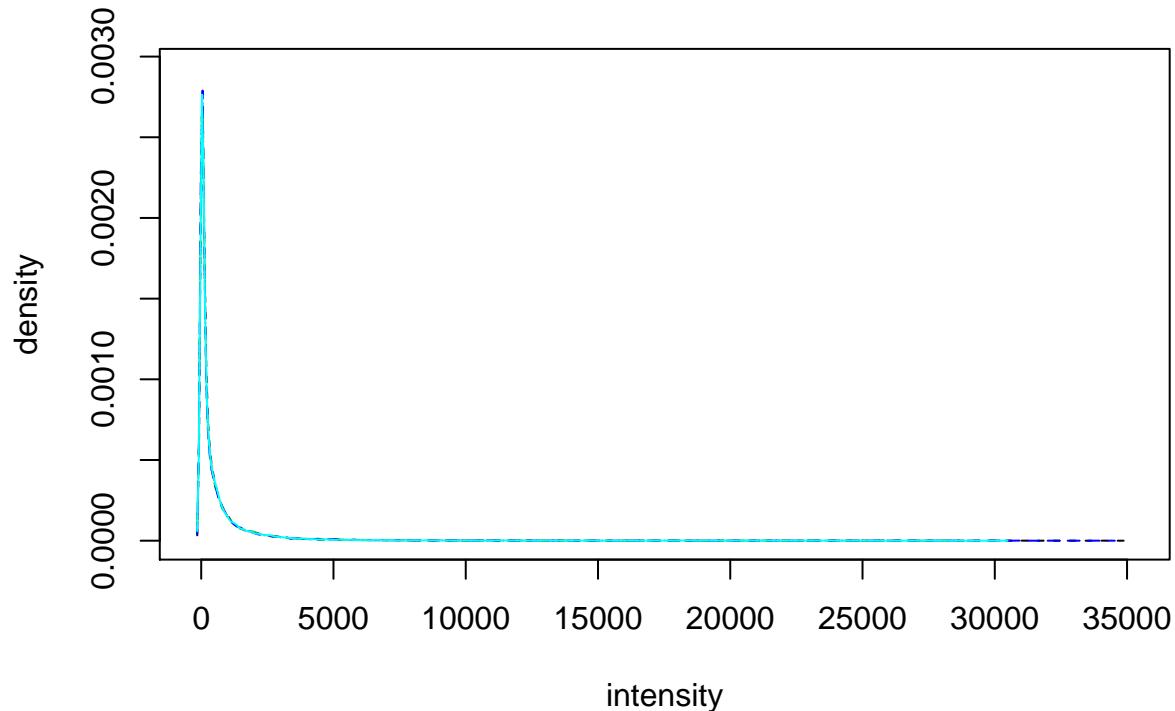
```
boxplot(eset.mas5)
```



Histogram with MAS5 normalization

```
hist(eset.mas5,col=0:6,main="Histogram of probe intensities")
```

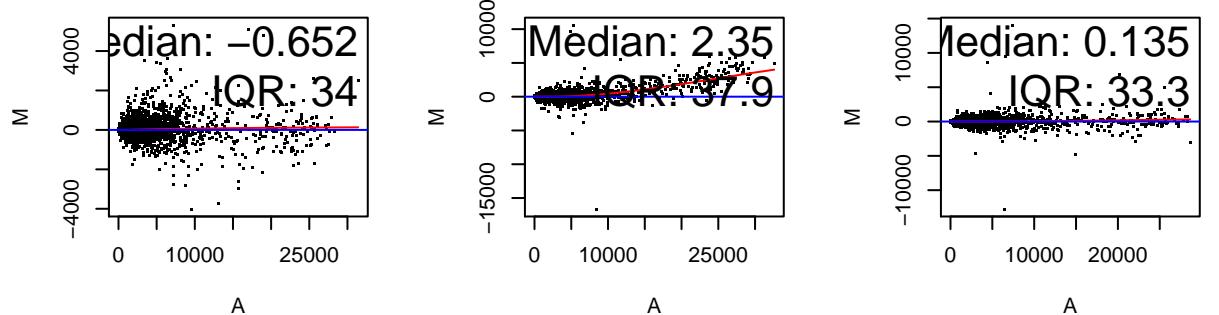
### Histogram of probe intensities



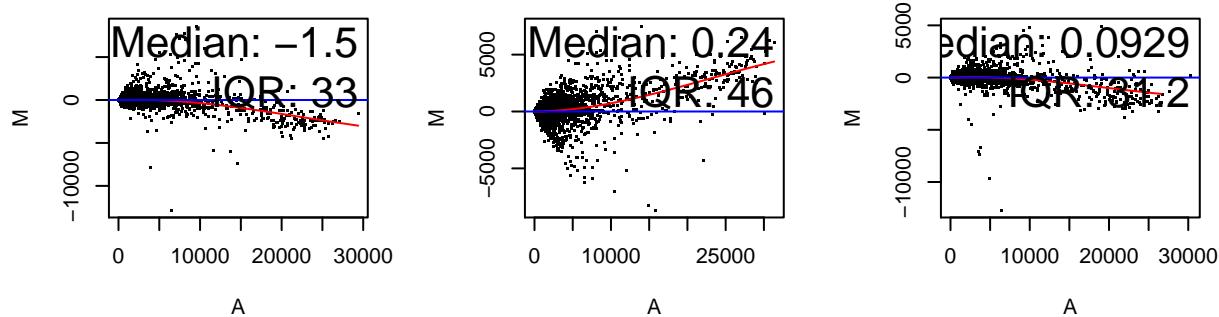
## MAplot with MAS5 normalization

```
par(mfrow=c(2,3))
MAplot(eset.mas5)
```

1.CEL vs pseudo-median reference.2.CEL vs pseudo-median reference3.CEL vs pseudo-median reference



1.CEL vs pseudo-median reference.2.CEL vs pseudo-median reference.3.CEL vs pseudo-median reference

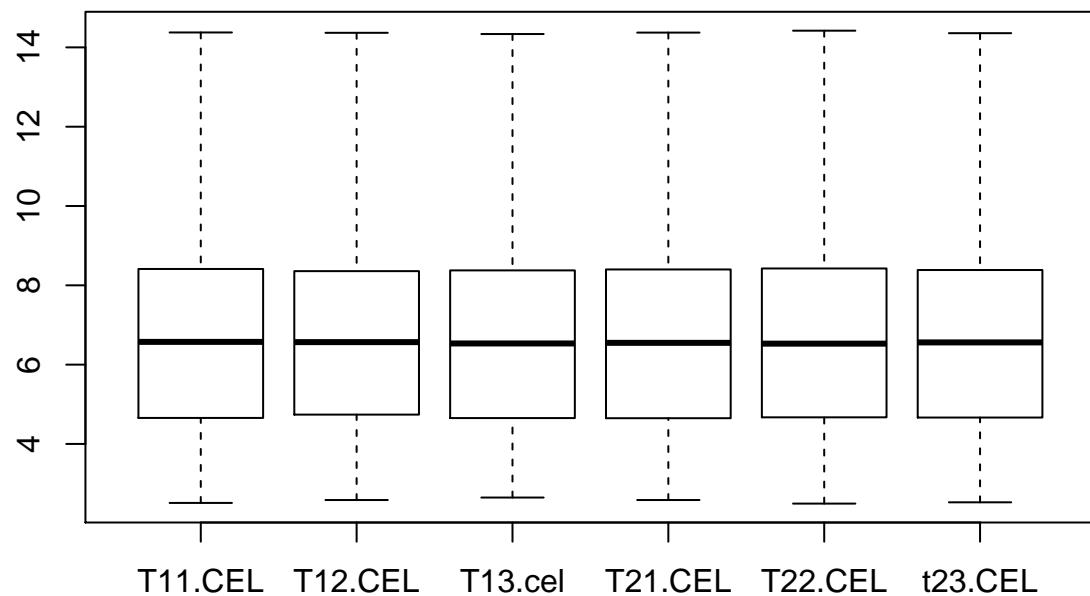


## Boxplot with RMA normalization

```
eset.rma=rma(a.data)
```

```
## Background correcting
## Normalizing
## Calculating Expression
```

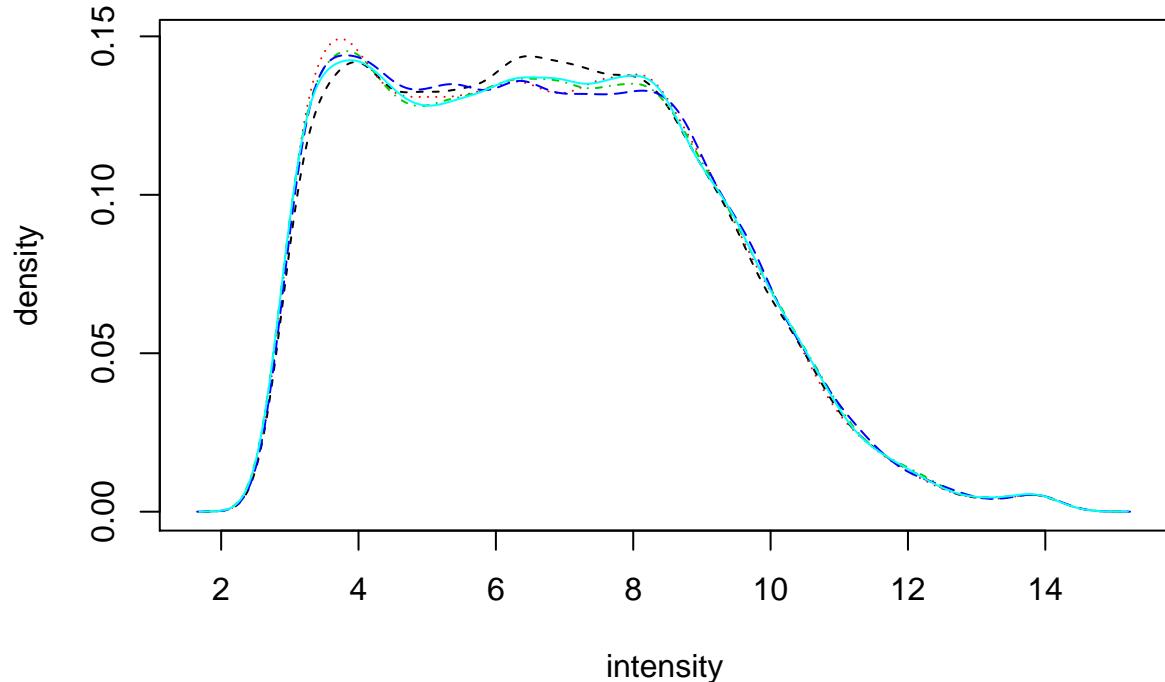
```
boxplot(eset.rma)
```



Histogram with RMA normalization

```
hist(eset.rma,col=0:6,main="Histogram of probe intensities")
```

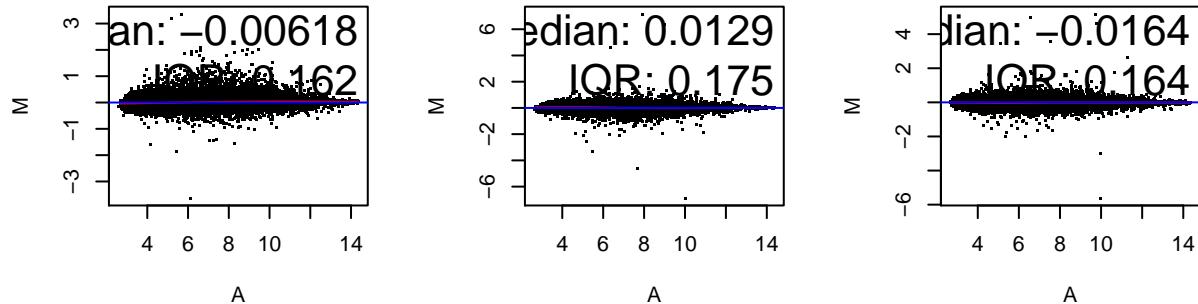
### Histogram of probe intensities



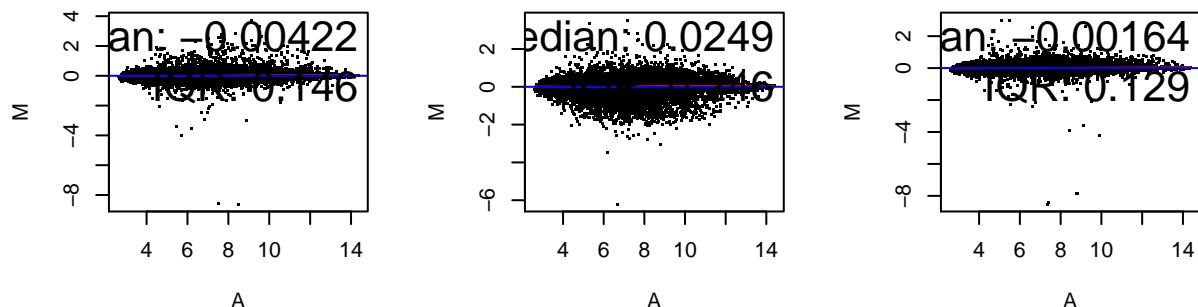
## MAplot with RMA normalization

```
par(mfrow=c(2,3))
MAplot(eset.rma)
```

1.CEL vs pseudo-median reference.CEL vs pseudo-median reference3.CEL vs pseudo-median reference



1.CEL vs pseudo-median reference.CEL vs pseudo-median reference1.CEL vs pseudo-median reference



## Question 2 From Homework 1. Tukey Biweight method

```
df = read.csv("/Users/muhtasim/Desktop/HW1/hw1.csv")

y= log2(df$PM1) - log2(df$MM1)

md<-median(y)

z<-abs(y-md)

mz<-median(z)

u<-(y-md)/(5*mz+.0001)
w<-ifelse(abs(u)<=1,(1-u^2)^2,0)
table.data=cbind(y,md,z,mz,u,w)
tuk_avg=sum(y*w)/sum(w)
tuk_avg
```

```
## [1] 2.921347
```

## Signal

```
IM<-ifelse(df$MM1>df$PM1,df$PM1/(2^tuk_avg),df$MM1)
i1<-IM
ys<-log2(df$PM1-i1)
ms<-median(ys)
zs<-abs(ys-ms)
ss<-median(zs)
us<-(ys-ms)/(5*ss+.0001)
ws<-ifelse(abs(us)<=1,(1-us^2)^2,0)
tbs<-sum(ws*ys)/sum(ws)
sgnl<-2^tbs
sgnl

## [1] 9727.687
```