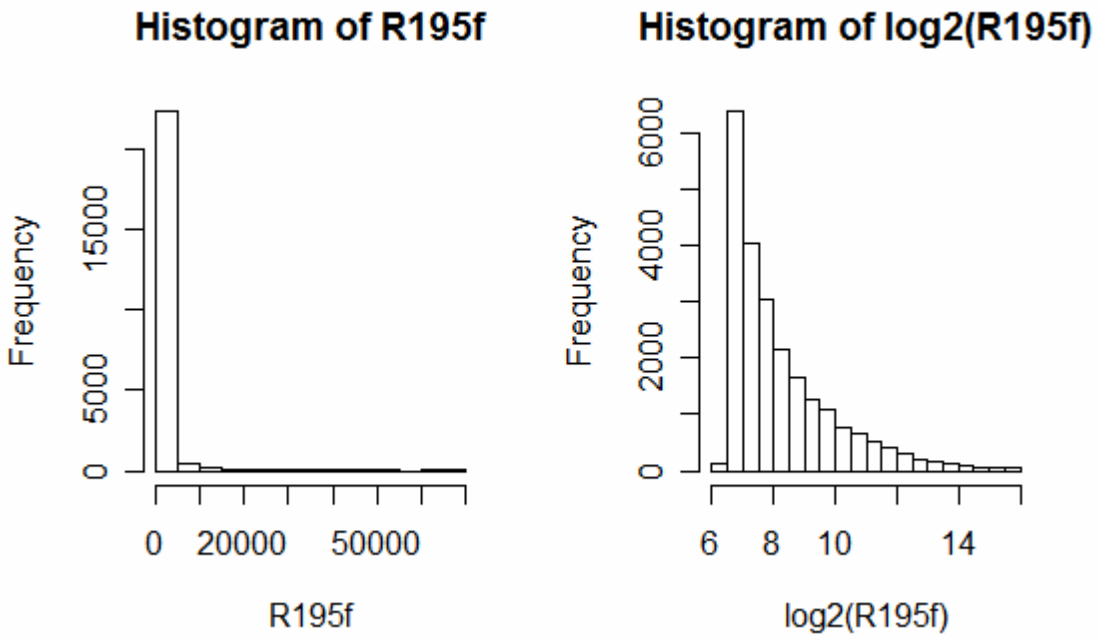
5.5 Representación gráfica de datos de experimentos de microarreglos. Algunos gráficos "a mano"

5.5.1 Histogramas

Seleccionamos los valores foreground de los canales Rojo y Verde del primer arreglo
Ya hemos visto que los datos de 6Hs.195.1.gpr están en la primera columna de beta7

```
R195f <- slot(beta7,"maRf")[,1] #primera columna de la matriz maRf
G195f <- slot(beta7,"maGf")[,1]#primera columna de la matriz maGf
```

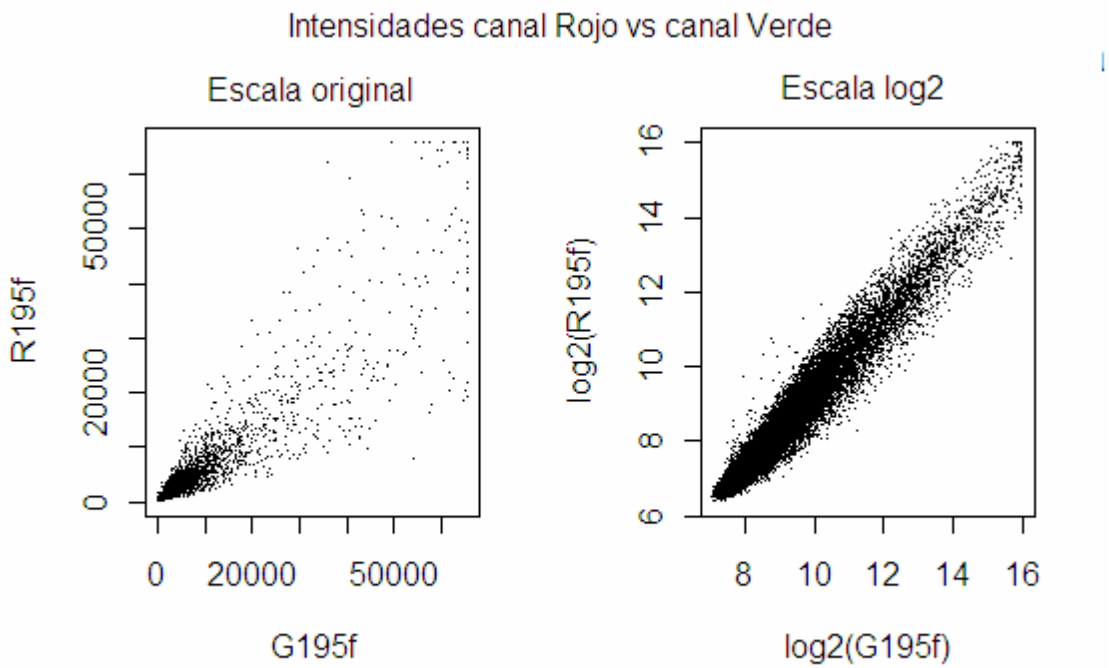
Vemos la distribución de las intensidades del canal rojo del primer arreglo en escala original y en escala logarítmica en base 2



Obtuvimos los gráficos en R con

```
> par(mfrow=c(1,2))
> hist(R195f)
> hist(log2(R195f))
```

5.5.2 Diagramas de Dispersión




```
> par(mfrow=c(1,2))
> plot(G195f,R195f, pch=".")
> plot(log2(G195f),log2(R195f), ",pch=".")
```

5.5.3 MAplot sin corrección por background

El gráfico más común es el denominado MA-plot. Se grafican los valores M (también conocido como log-ratio) en el eje vertical y los valores A en el horizontal.
¿Cómo se calculan dichos valores?
Se definen los valores M y A de la siguiente manera.

$$M = \log_2 \frac{R_f}{G_f}$$

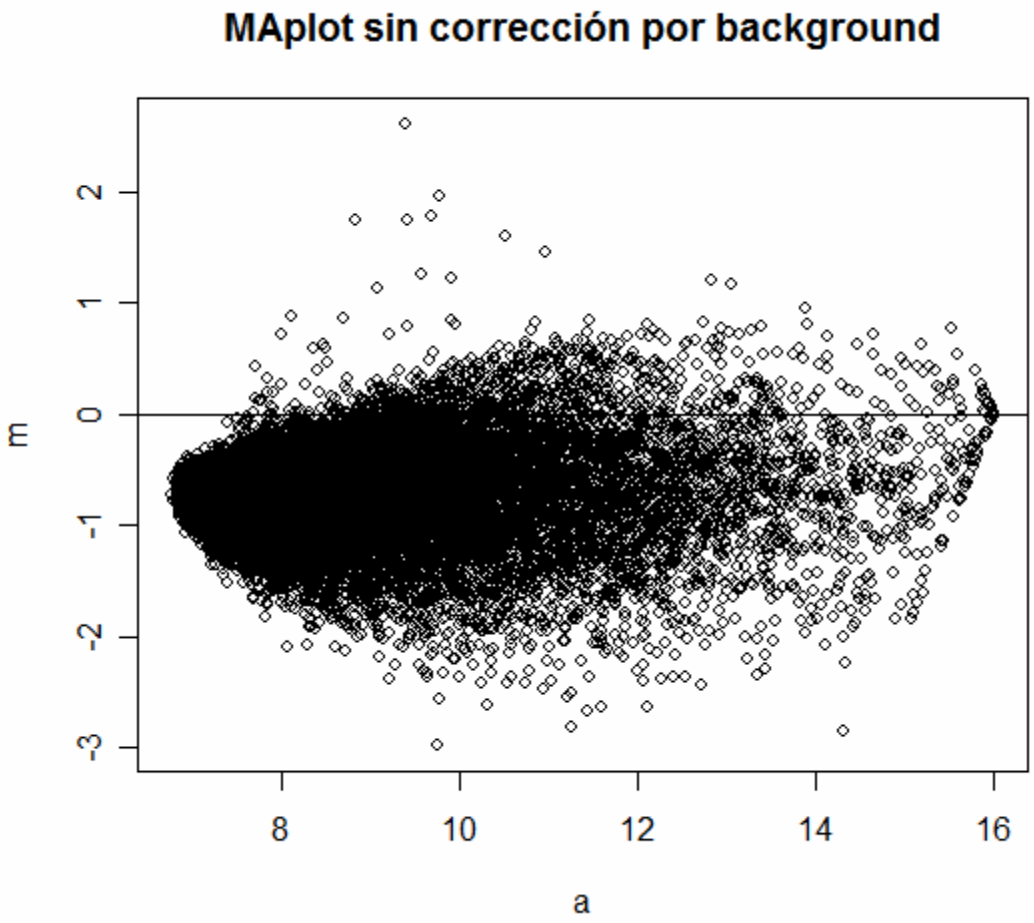
$$A = \frac{\log_2 R_f + \log_2 G_f}{2}$$

Para construir un gráfico MA-plot, para el primer arreglo de beta7, en R ingresamos los siguientes comandos

```
> m <- log2(R195f/G195f)
> a <- 0.5*log2(R195f*G195f)

> par(mfrow=c(1,1))
> plot(a,m, main="MAplot sin corrección por background")
> abline(0,0)
```

y obtenemos

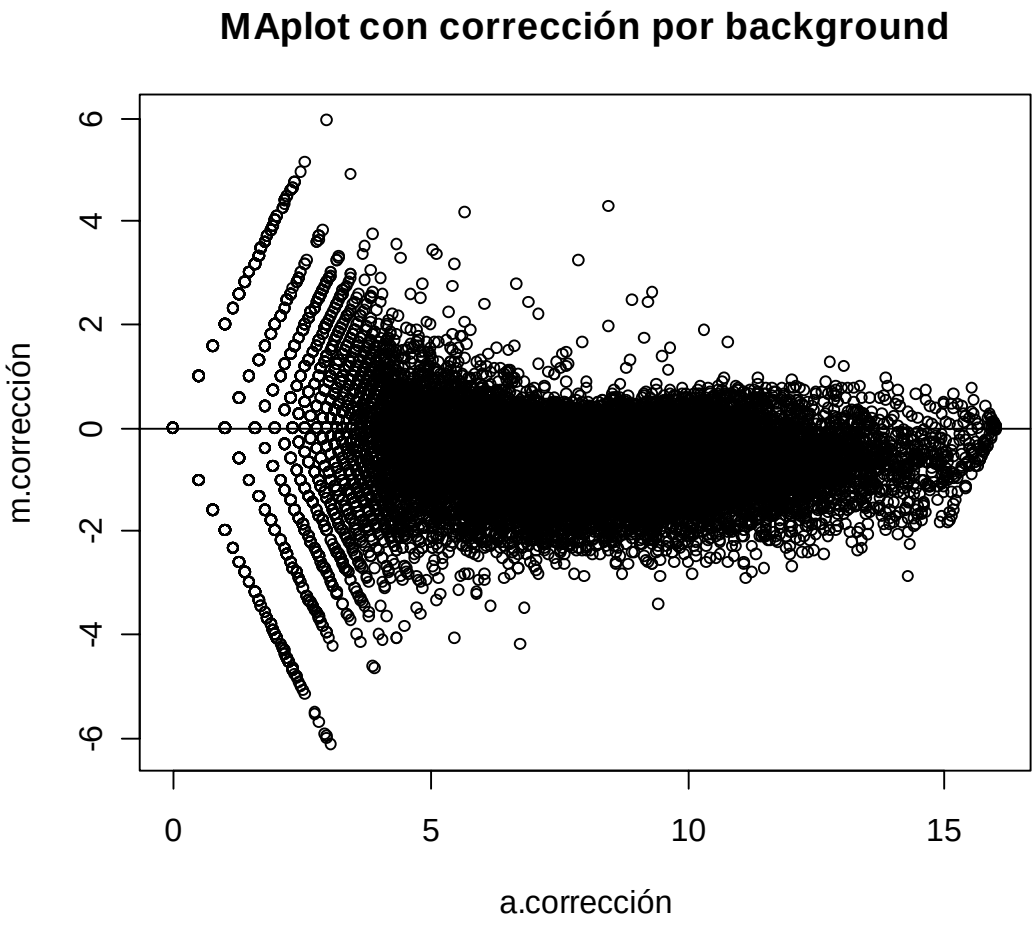


Un gráfico MA-Plot ideal es aquel en el cual toda la nube de puntos se distribuye en forma pareja alrededor del 0 del eje M.

Todas aquellas desviaciones con respecto al ideal se deben principalmente a errores sistemáticos vinculados con varios factores, como ser: diferencias en la eficiencia de la incorporación de los tintes, diferencias en la cantidad de mRNA en las muestras, diferencias en los parámetros de escaneado, diferencias entre los grupos de agujas, efectos espaciales y efectos debido al plato.

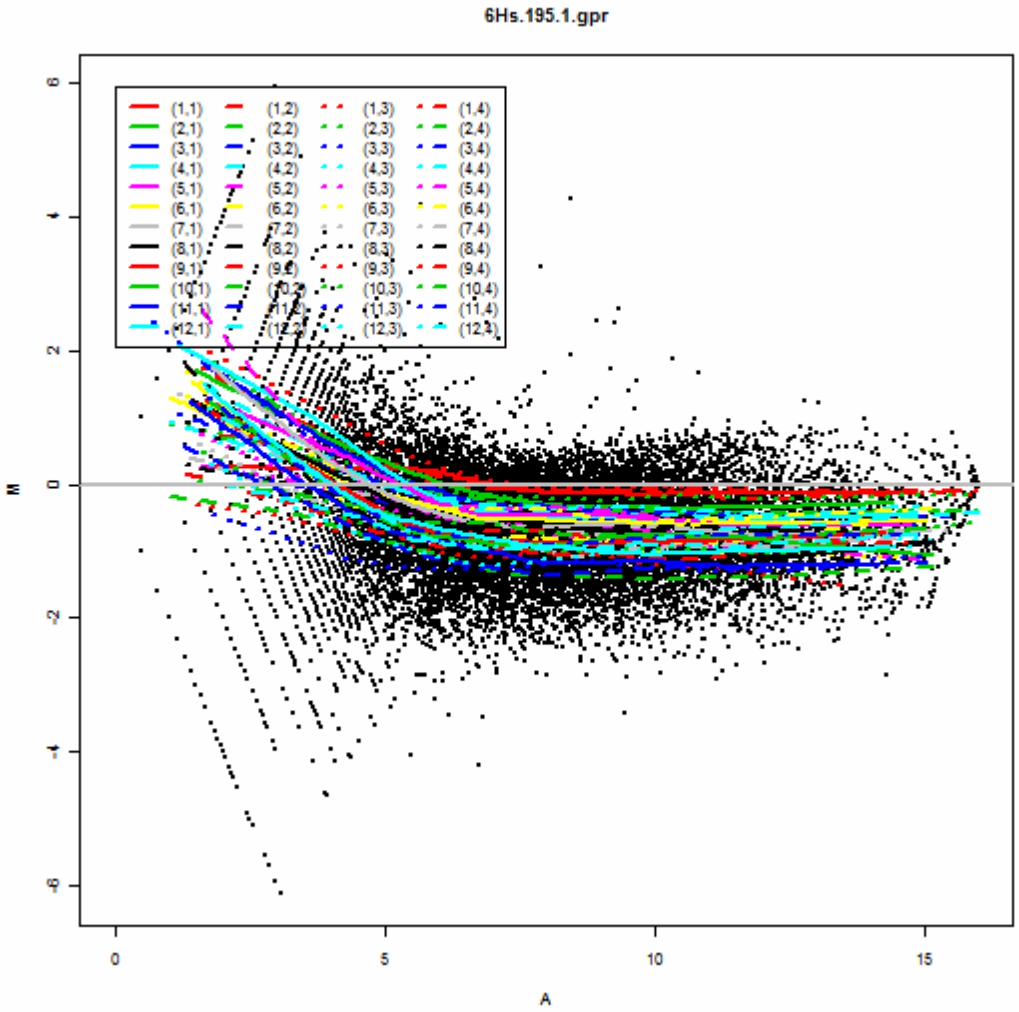
5.5.3 MAplot con corrección por background

```
> par(mfrow=c(1,1))
> m.corrección <- log2((R195f-R195b)/(G195f-G195b))
Warning message:
NaNs produced in: log(x, base)
> a.corrección <- 0.5*log2((R195f-R195b)*(G195f-G195b))
Warning message:
NaNs produced in: log(x, base)
>
>
> plot(a.corrección,m.corrección, main="MAplot con corrección por background")
> abline(0,0)
```



5.5.5 Función `maPlot`

```
X11( pointsize =7)
maPlot(beta7[,1])
```



¿Qué son las curvas que aparecen por defecto cuando se realiza un maPlot?

Son curvas Print tip Loess

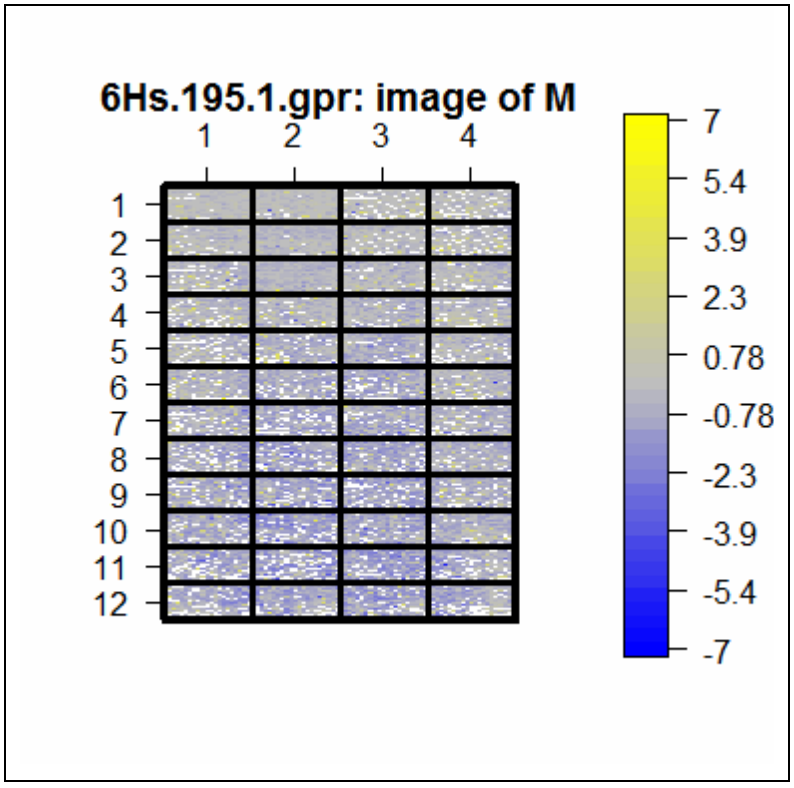
Veremos algo más sobre estas curvas en la próxima clase.

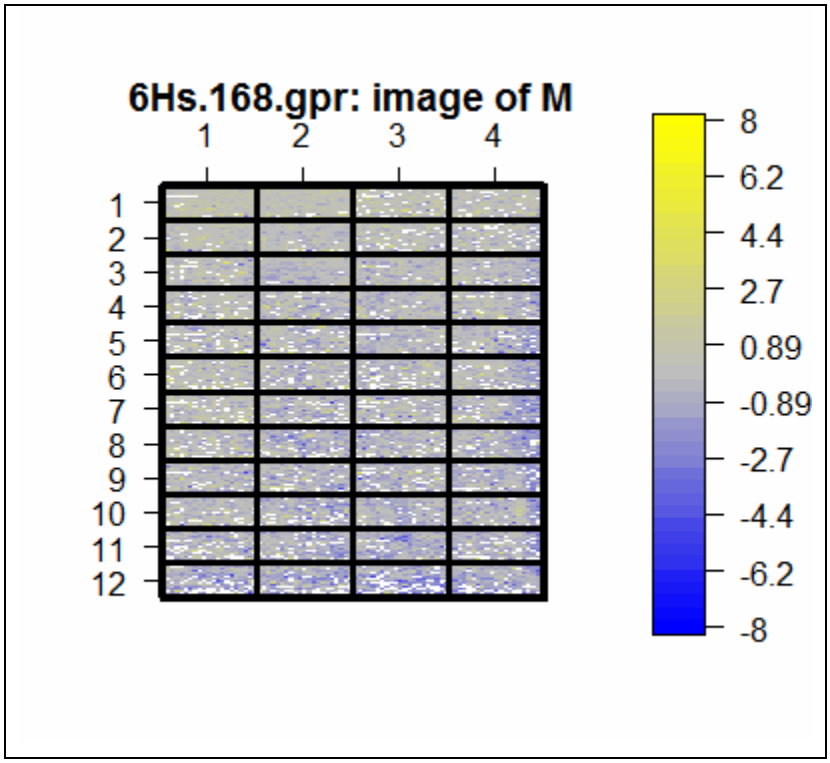
5.6 Gráficos espaciales

Los gráficos espaciales son gráficos de dos dimensiones en el cual cada punto representa un estadístico del spot del microarreglo. Se obtienen mediante la función **`image`** y se utilizan para explorar la calidad de hibridación de cada arreglo.

Los siguientes comandos permiten obtener los gráficos espaciales para los valores de M de los dos primeros arreglos de beta7, con X11() se abre una ventana para cada gráfico:

```
for(i in 1:2) {
   X11()
   image(beta7[,i])
}
```

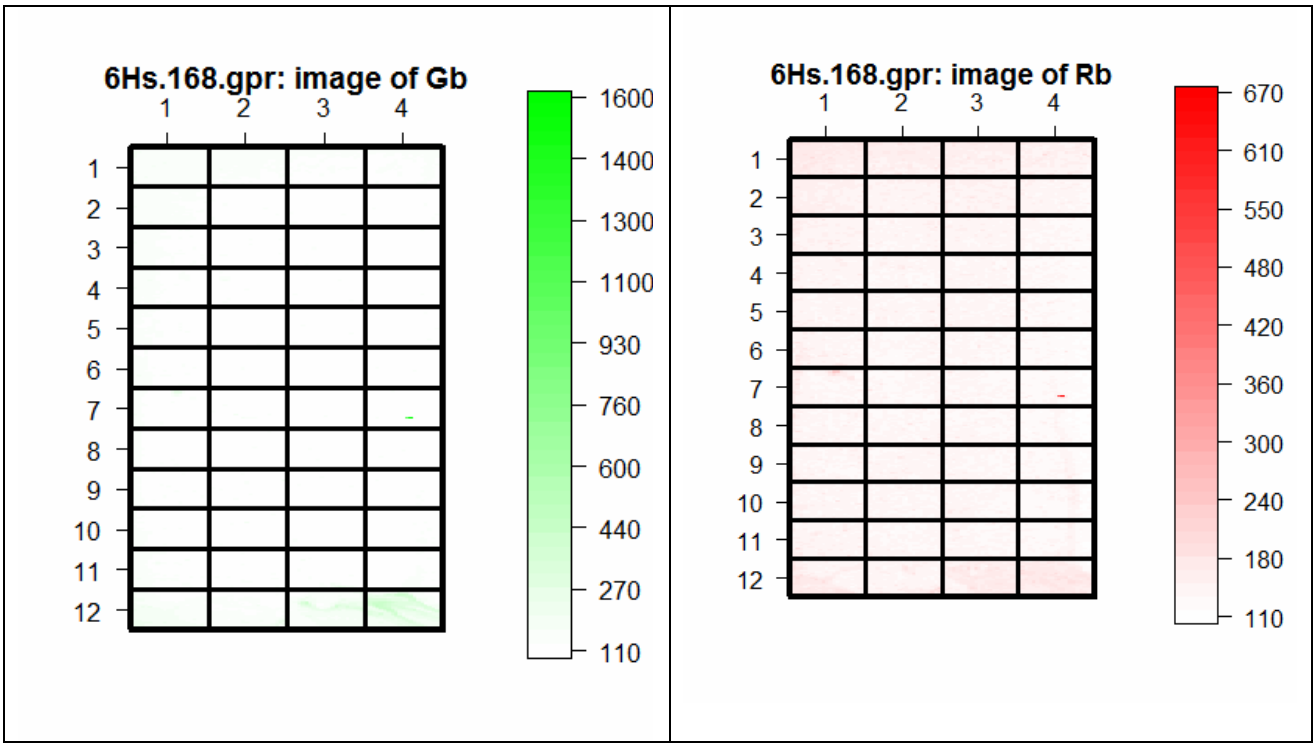




Por defecto la función **`image`** produce un gráfico espacial de M. Es posible modificar este valor por defecto utilizando el parámetro `xvar` Si se utiliza `xvar ="maRb"` se obtiene la el gráfico espacial para el background para el canal Cy5 (rojo). Por defecto se despliega la barra de escala de colores, **`bar=TRUE`**, que se puede eliminar con **`bar= FALSE.`**

```
for(i in 1:2) {
   X11()
   image(beta7[,i],xvar="maRb")
}


for(i in 1:2) {
   X11()
   image(beta7[,i],xvar="maGb")
}
```



5.7 Boxplots

El boxplot, es un gráfico propuesto por Tukey en 1977. Está basado en medidas resumen de la distribución de un conjunto de datos. Las medidas resumen son: la mediana, los cuartiles superior e inferior, el rango y posiblemente algunos valores extremos. La caja central del gráfico representa el *rango intercuartil* (**IQR**), está determinada por el percentil del 25% y el del 75%, contiene el 50% central de los datos. La línea dentro de la caja representa a la mediana, una medida de posición del centro de los datos. Los valores que se encuentran más allá de **1.5 IQR** por encima del percentil del 75% y por debajo del percentil del 25% se grafican individualmente.

La función **`boxplot`** produce boxplots para objetos de clase **`marrayRaw`** y **`marrayNorm`** que veremos próximamente.

Los argumentos principales de la función **`boxplot`** son:

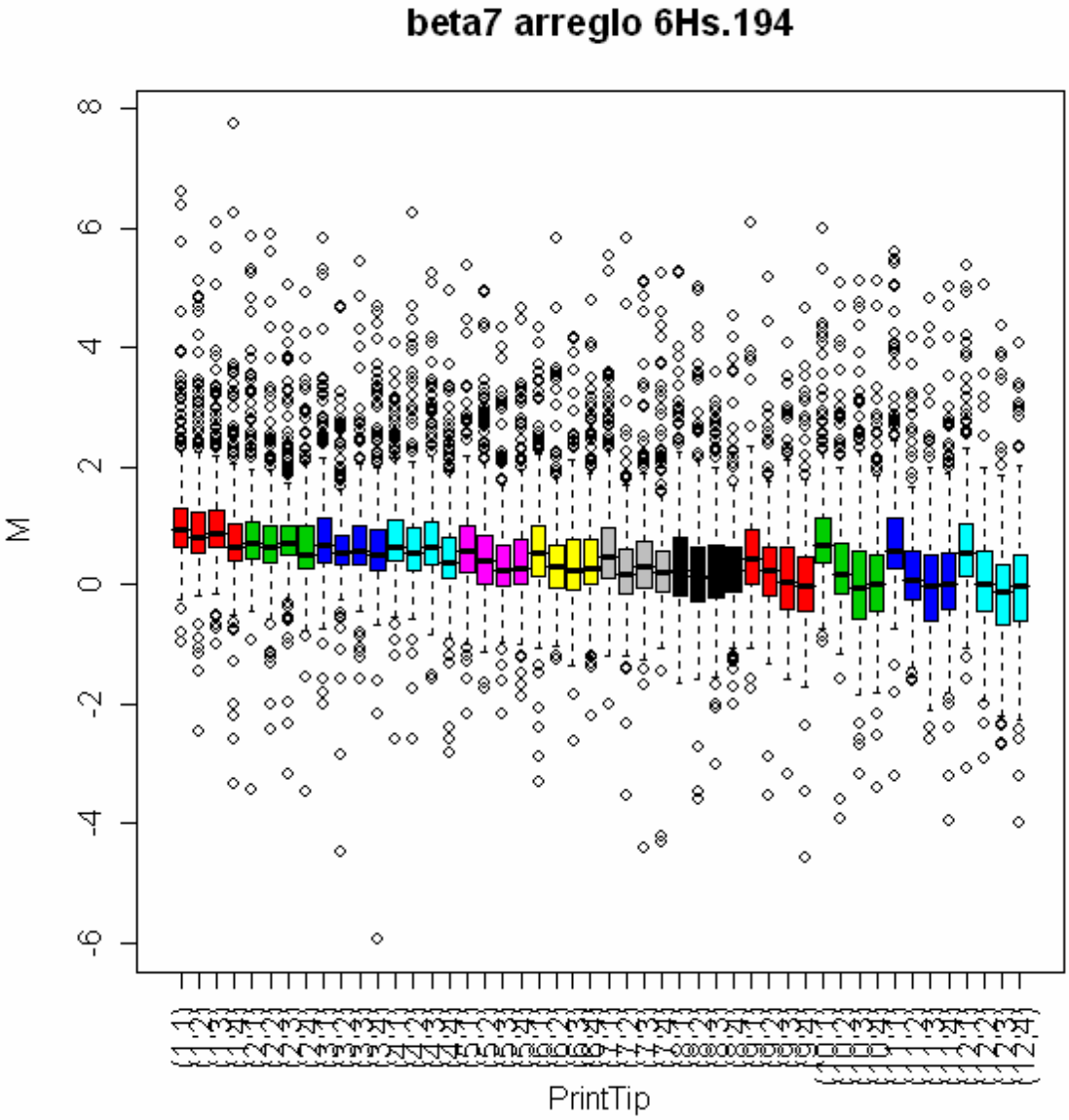
```
x: Objeto de clase marrayRaw o  marrayNorm.
xvar: Nombre del método para estratificar los datos. Por ejemplo el
      nombre de un slot de un objeto “layout” como maPlate o un  método
      como maPrintTip. Si xvar = NULL, no se estratifican los    datos.
yvar: Nombre del método para la obtención del estadístico de interés  por
```

ejemplo **maM**.

Con

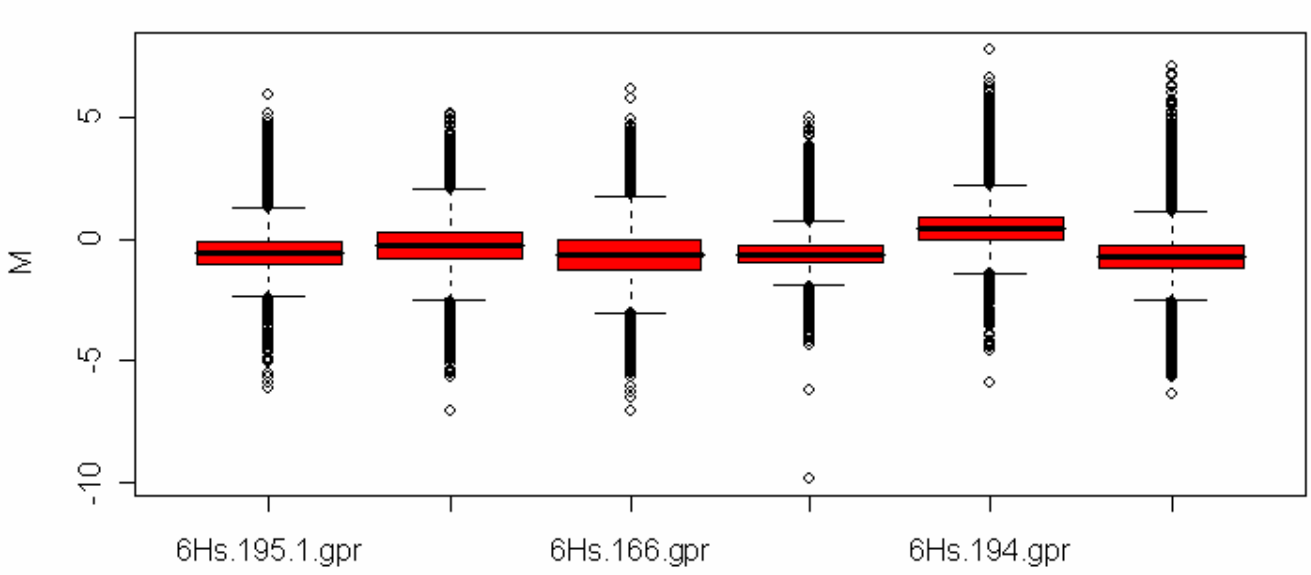
> **boxplot(beta7[, 5], xvar = "maPrintTip", yvar = "maM",main="beta7 arreglo 6Hs.194")**

obtenemos



Vemos una tendencia decreciente en los cocientes (ratios) de intensidades en relación al print-tip group en el arreglo 5. Los valores correspondientes a las agujas de la primera fila son superiores a los de la última. Este sesgo se corregirá mediante la normalización por grupo de aguja.
Por defecto la función boxplot produce un boxplot para cada arreglo:

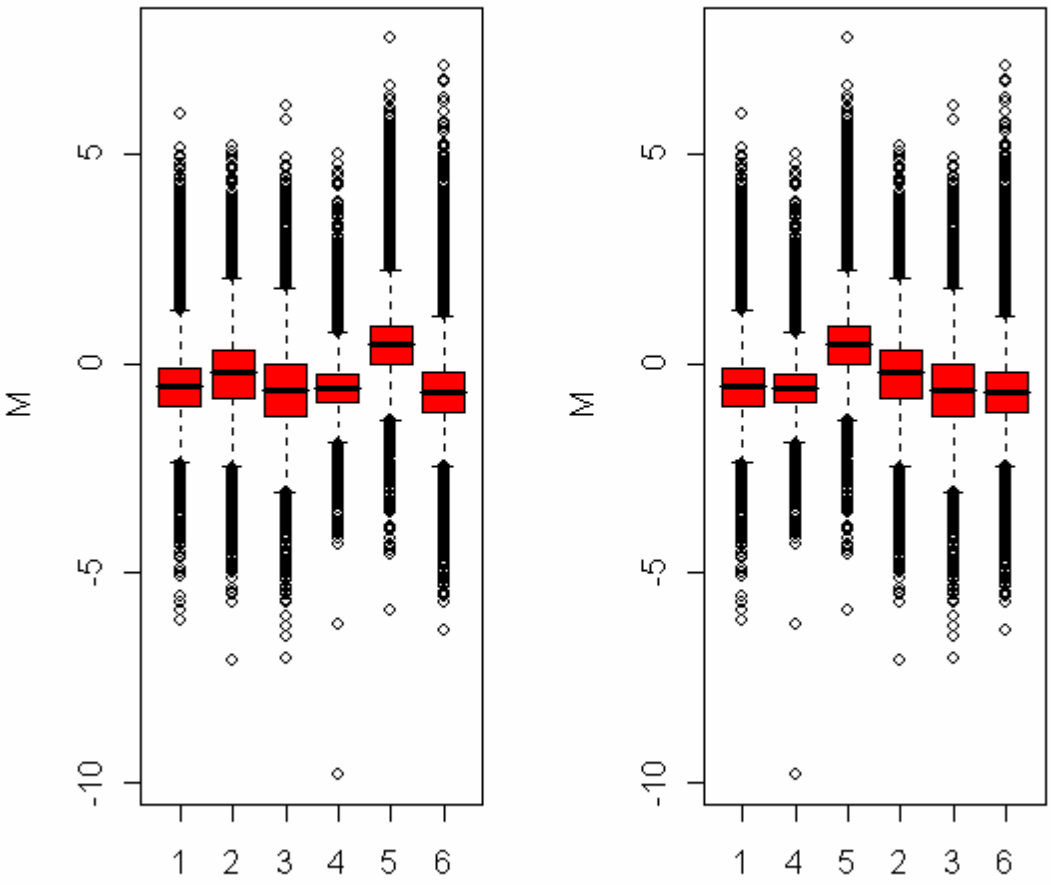
```
> boxplot(beta7,yvar="maM")
```



Recordemos

```
@maInfo
      FileNames SubjectID  Cy3  Cy5 Date of Blood Draw
1 6Hs.195.1.gpr         1 b7 - b7 +         2002.10.11
2   6Hs.168.gpr         3 b7 + b7 -         2003.01.16
3   6Hs.166.gpr         4 b7 + b7 -         2003.01.16
4 6Hs.187.1.gpr         6 b7 - b7 +         2002.09.16
5   6Hs.194.gpr         8 b7 - b7 +         2002.09.18
6 6Hs.243.1.gpr        11 b7 + b7 -         2003.01.13


> par(mfrow=c(1,2))
> boxplot(beta7,yvar="maM",names=c("1","2","3","4","5","6"))
> boxplot(beta7[,c(1,4,5,2,3,6)], yvar = "maM",
names=c("1","4","5","2","3","6"))
```


M
5
0
-5
-10
1
2
3
4
5
6
M
5
0
-5
-10
1
4
5
2
3
6