

# 1 Métodos Computacionales para Calidad y $6\sigma$

## Resumen

En el presente trabajo se ensaya distintos métodos del software de estadística R-cran para diseñar, implementar, controlar y mejorar sistemas de calidad basados en las recomendaciones de Seis Sigma. En este capítulo trabajaremos con Gráficos con R.

## Palabras Clave:

Calidad, Seis Sigma, Estadística, Gráficos vectoriales.

## 2 Introducción

Aquí nos centraremos en describir el contexto de aplicación de la solución que pretendemos presentar en el trabajo de gráficos con R. Los gráficos son particularmente importantes para los proyectos Seis Sigma. El objetivo de un gráfico usualmente es ayudar en la interpretación de datos. Por consiguiente, proveer una adecuada explicación de los datos a través de gráficos es crucial. Gráficos de dos dimensiones pueden ser usados como la base para representaciones más complejas, y datos multidimensionales pueden ser mostrados a través de extensiones de dichos gráficos. Antes de describir los principales gráficos usados en los proyectos de Seis Sigma, en esta sección vamos a explicar algunos conceptos importantes respecto a los gráficos en general.

### 2.1 Uso de gráficos

Los gráficos son muy útiles en las primeras etapas de análisis, cuando necesitamos una descripción del proceso para hacer conjeturas, descifrar relaciones, y planear investigaciones posteriores. Así, el análisis descriptivo es la misión más importante de los gráficos. En el último paso de nuestro análisis, cuando presentamos nuestros resultados, los gráficos serán probablemente la tarjeta de negocio de nuestras conclusiones. Por lo tanto, elegir el tipo adecuado de gráfico para comunicar nuestros avances o propuestas es crucial.

### 2.2 Conceptos de Fondo

La mayoría de los gráficos que usamos en estadística (y por consiguiente en Seis Sigma) son gráficos bidimensionales. La naturaleza de las variables que estamos midiendo en cada dimensión determina la escala del gráfico. Otra decisión sobre nuestro gráfico es establecer sus límites.

**Ejemplo 1.** Cartucho para impresora La siguiente tabla muestra los datos de una muestra de 24 cartuchos de impresora de una marca determinada. Son de tres diferentes llenadores, y hay dos tipos de cartucho: color y negro. El

volumen y la densidad en cada cartucho se midieron. El conjunto de datos `ss.data.pc` en el paquete `SixSigma` contiene estos datos.

**Recuerde instalar y cargar la biblioteca**

- `install.package("SixSigma")` solo una vez.
- `library(SixSigma)` cada vez que escribo un script nuevo.

```
> library(SixSigma)
> listado <- data(package="SixSigma")
> # listado$results
> listado$results[13, ]
```

Package
"SixSigma"
LibPath
"/home/rpalma/R/x86_64-pc-linux-gnu-library/3.4"
Item
"ss.data.pc"
Title
"Data set for the printer cartridge example"

Mostraremos la cabecera de la tabla de datos a utilizar.

```
> head(ss.data.pc)
```

	pc.col	pc.filler	pc.volume	pc.density	pc.batch	pc.op
1	C	1	16.75331	1.253419	1	A
2	C	2	18.01435	1.112435	1	B
3	C	3	15.64487	1.141105	1	C
4	C	1	18.02817	1.091772	1	D
5	C	2	13.78313	1.146547	2	A
6	C	3	16.75840	1.123339	2	B

De manera que ahora, se mostrará la relación entre la densidad y el volumen de los datos con los ajustes ya definidos para la función `plot` por defecto.

```
> pc.volume <- ss.data.pc$pc.volume
> pc.density <- ss.data.pc$pc.density
> plot (pc.volume ~ pc.density, pch = 16)
```

### 3 Gráfico de Barras

Un gráfico de barras es un gráfico muy simple donde se muestran algunas cantidades como la altura de las barras. Cada barra representa un factor donde la variable en estudio se está midiendo. Un gráfico de barras es generalmente

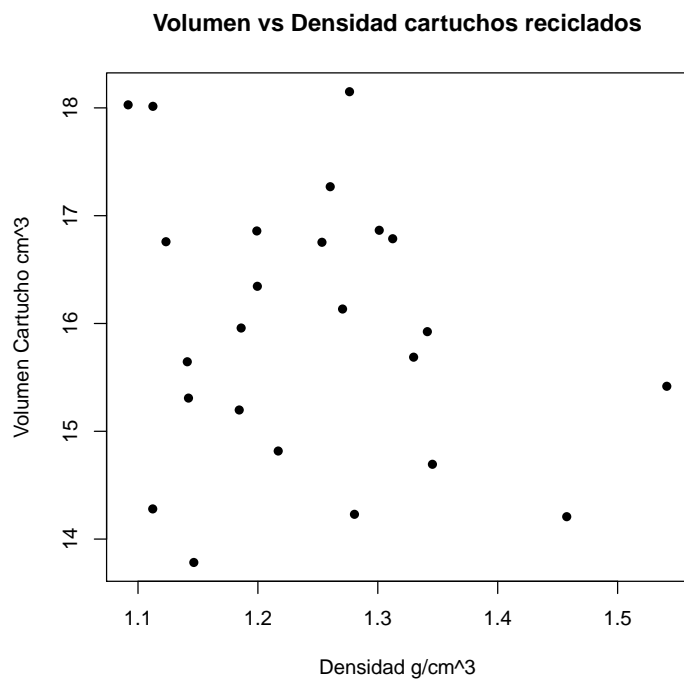


Figure 1: Ejemplo Gráfica Densidad Volúmen de Cartuchos Toner

la mejor representación gráfica para el recuento. **Ejemplo 2.** El fabricante de cartucho de impresora distribuye su producto a cinco regiones. Una cantidad inesperada de cartuchos defectuosos ha sido devuelta en el último mes. El diagrama de barras en la figura 2 es una manera directa para ver los datos de la Tabla 2 (disponible en el paquete SixSigma dentro del conjunto de datos `ss.data.pc.r`). Utilizamos el siguiente código en R a partir de la función `barplot`, tomando en cuenta que podemos hacer un gráfico de barras para cada cartucho y cambiando los colores.

```
> head(ss.data.pc.r)

  pc.regions pc.def.a pc.def.b pc.def
1  region.1      37      63    100
2  region.2      17      48     65
3  region.3      38      39     77
4  region.4      41      43     84
5  region.5      13      54     67
```

## 4 Histogramas

Un histograma es un gráfico de barras para las variables continuas. Este gráfico de barras muestra la distribución de las mediciones de las variables. En el eje x, cada barra representa un intervalo de los valores posibles de una variable. La altura de las barras (es decir, el eje y) depende de la frecuencia (relativa o absoluta) de las medidas dentro de cada intervalo. La regla es que el área de las barras debe ser proporcional a las frecuencias.

En el **Ejemplo 3**, tomaremos los datos del primer ejemplo que realizamos y construiremos un gráfico de barras de volumen vs densidad con la función `hist` (la gráfica incluye las líneas de densidad).

En este histograma, hemos cambiado el color de las barras. La línea continua es una línea de densidad para los datos. La línea discontinua es la línea de densidad teórico para una distribución normal. Los parámetros de esta distribución teórica han sido anotados en el gráfico.

## 5 Diagrama de Dispersión

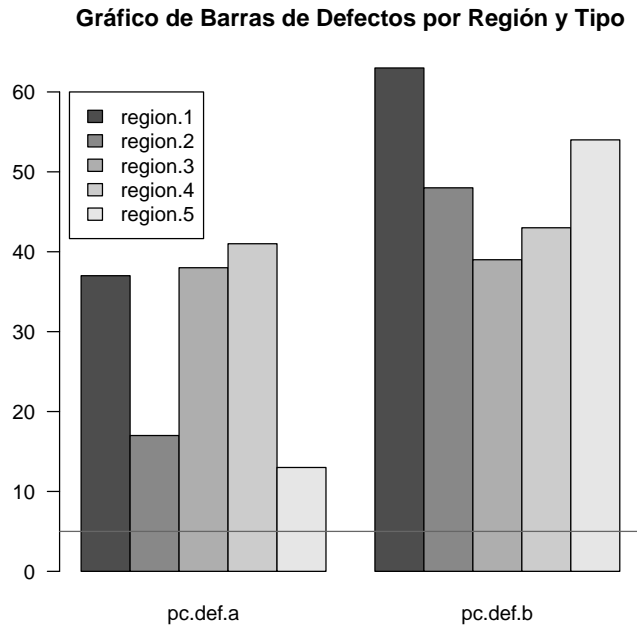
Un diagrama de dispersión es una herramienta importante para revelar las relaciones entre dos variables. En lenguaje estadístico, estas relaciones pueden ser subsumidas en el concepto de correlación. Por lo tanto, podemos tener tres tipos de correlación entre dos variables:

- La correlación positiva: altos valores de una de las variables conducen a valores altos de la otra.
- Correlación negativa: valores altos de una de las variables conducen a valores bajos de la otra.

```

> barplot(as.matrix(ss.data.pc.r[,2:3]),
+ las = 1,
+ beside = TRUE,
+ legend = ss.data.pc.r[,1],
+ args.legend = list(x=3.5,y=60),
+ main = "Gráfico de Barras de Defectos por Región y Tipo",
+ sub = "Ejemplo 2 Cartuchos Reciclados")
> abline(h = 5,
+ col = "#666666")

```



Ejemplo 2 Cartuchos Reciclados

Figure 2: Ejemplo monocromo: Siempre es preferible hacer informes en ByN y evitar el color

```

> hist (ss.data.pc$pc.volume,
+ main = "Printer Cartridge Volume",
+ xlab = "Volumen", ylab="Densidad",
+ col = "#BBBBBB",
+ border = "white",
+ bg = "red",
+ freq = FALSE,
+ ylim = c(0,0.4))
> curve (dnorm(x,16,1),
+ add = TRUE,
+ lty = 2,
+ lwd = 2)
> lines (density(ss.data.pc$pc.volume),
+ lwd = 2,col="red")
> text (label = expression(paste(mu==16,
+ "; ",
+ sigma==1,
+ sep = "")),
+ x = 16.5,
+ y = 0.4,
+ adj = c(0,1))
> grid()
> box()
>
>
>

```

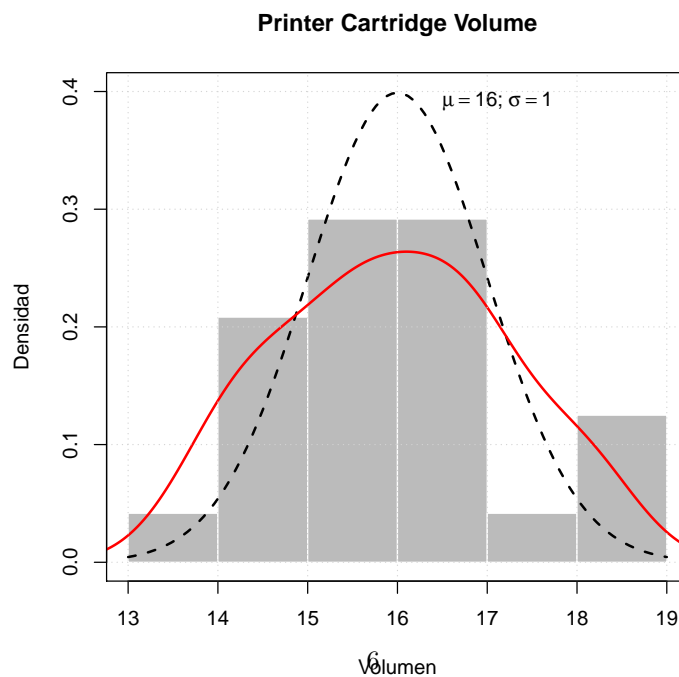


Figure 3: Histograma con líneas densidad (ejemplo cartucho).

- No existe correlación: las variables son independientes.

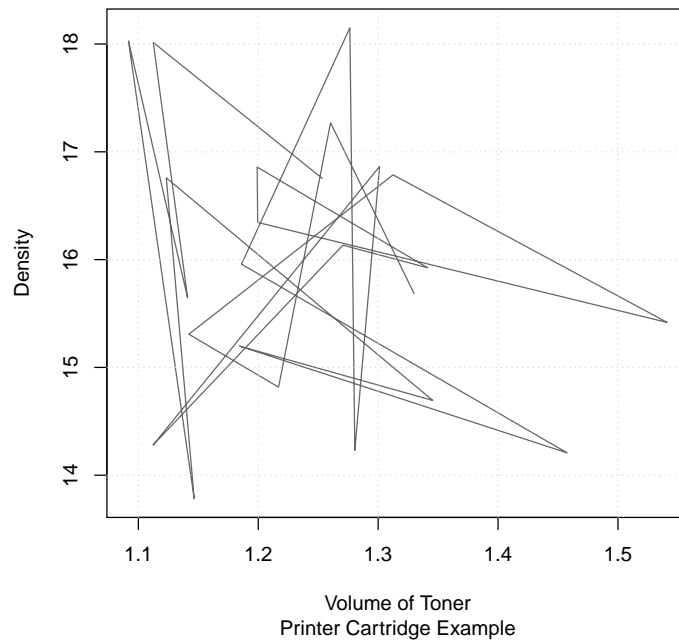
Esta técnica es la base del análisis multidimensional multivariado y es el corazón del marco teórico con el que trabajó Taguchi en los inicios de estandar ISO 9000

## 6 Modelo de Regresión

En el conjunto de datos `ss.data.pc`, tenemos dos variables continuas: `pc.volume` y `pc.density`. Si queremos comprobar si la densidad y el volumen están relacionados, la primera cosa que tenemos que hacer es generar un diagrama de dispersión para encontrar patrones de esta relación. En el **Ejemplo 4**, utilizamos el siguiente código para producir el diagrama de dispersión en la figura 4. En este caso volvemos a utilizar nuevamente la función `plot`, pero con la variante de línea en lugar de puntos.

```
> pc.volume <- ss.data.pc$pc.volume
> pc.density <- ss.data.pc$pc.density
> plot (pc.volume ~ pc.density,
+ main = "Buscando Correlación entre Densidad y Volumen de Imrpimible", type="l",
+ col = "#666666",
+ pch = 16,
+ sub = "Printer Cartridge Example",
+ xlab = "Volume of Toner",
+ ylab = "Density")
> grid()
```

### Buscando Correlación entre Densidad y Volumen de Imprimibl

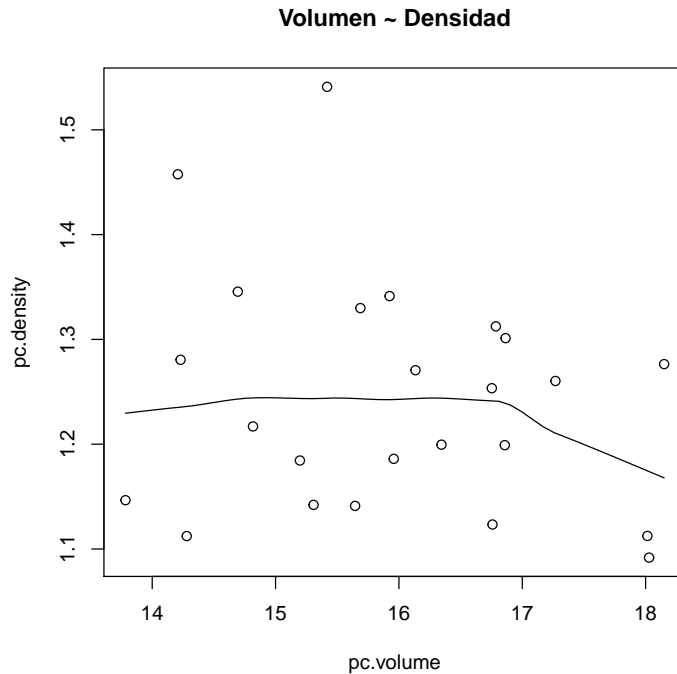


#### 6.0.1 Técnica de la Dispersión Suave

Esta técnica (smooth scatter en inglés) utiliza el criterio de suavizado tratando de minimizar el error cuadrático medio de a tramos de dos o tres pares ordenados por turno. A la línea obtenida se la suaviza con spline o bezier.

```
> scatter.smooth(x=pc.volume, y=pc.density, main="Volumen ~ Densidad") # scatterplot
```





### 6.0.2 Gráfico de Serie Tiempo

Un gráfico de serie es un gráfico bidimensional donde el eje x representa una línea de tiempo y en el eje de ordenadas se representa una variable que queremos monitorizar. Esta variable puede ser un parámetro críticos para la calidad (CTQ) y característica de nuestro proceso o por lo menos un parámetro que la afecta en alto grado.

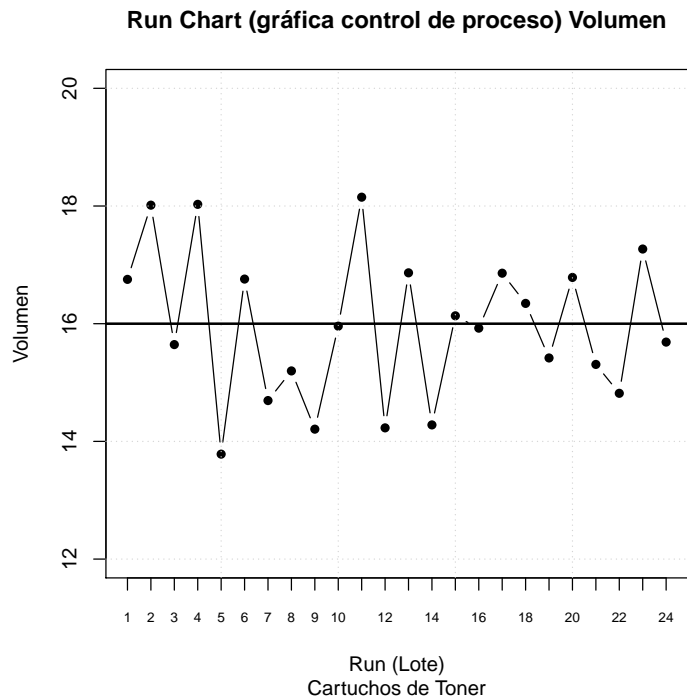
Los gráficos de serie nos permiten detectar patrones que pueden ser indicativos de cambios en un proceso. Los cambios implican variabilidad y, por tanto, menos calidad. En particular, si se detecta ciclos, tendencias o cambios, se debe revisar nuestro proceso. **Ejemplo 5.** Queremos controlar el volumen en los cartuchos, disponibles en el conjunto de datos `ss.data.pc`. Supongamos que el valor objetivo para el volumen es de 16. Usando el siguiente código obtenemos el gráfico de serie en la figura 5. Empleamos la función `plot` para realizar el ejemplo.

```
> plot (ss.data.pc$pc.volume,
+ type = "b",
+ pch = 16,
+ ylim = c(12,20),
+ axes = FALSE,
+ main = "Run Chart (gráfica control de proceso) Volumen",
```

```

+ sub = "Cartuchos de Toner",
+ xlab = "Run (Lote)",
+ ylab = "Volumen")
> axis(1,
+ at = 1:24,
+ cex.axis = 0.7)
> axis(2)
> box()
> grid()
> abline(h = 16,
+ lwd = 2)
>

```

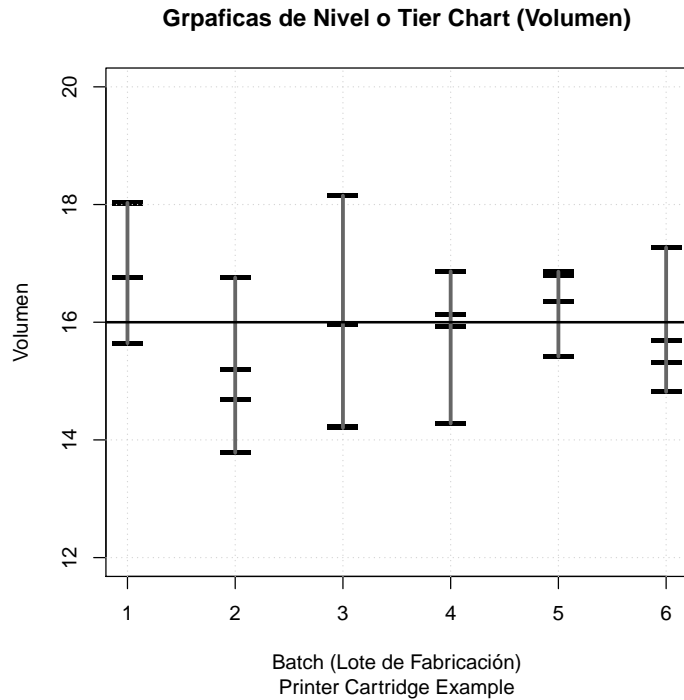


## 7 Gráfico de Nivel

Un gráfico de nivel es similar a un gráfico de serie. Utilizamos gráficos de nivel cuando tenemos más de una observación en cada serie (por ejemplo, lotes, días, etc.). Con el gráfico de nivel podemos ver variación a corto plazo y la variación a largo plazo de manera conjunta en un solo gráfico. A corto plazo la variación es una dentro de cada subgrupo, mientras que la variación a largo plazo es aquella entre todos los grupos.

**Ejemplo 6.** Supongamos que la muestra de 24 cartuchos proviene de 6 lotes diferentes y que se miden secuencialmente. Es decir, los cartuchos de 1 a 4 son del lote 1, 5 a 8 son del lote 2, y así sucesivamente. La información se encuentra en la variable `pc.batch`. Podemos crear el gráfico utilizando la función `Stripchart` y añadiendo una línea para cada subgrupo (Figura 6.):

```
> stripchart(pc.volume ~ pc.batch,
+ data = ss.data.pc,
+ pch = "-",
+ cex = 3,
+ xlab = "Batch (Lote de Fabricación)",
+ ylab = "Volumen",
+ ylim = c(12,20),
+ vertical = TRUE,
+ main = "Gráficas de Nivel o Tier Chart (Volumen)",
+ sub = "Printer Cartridge Example")
> grid()
> for (i in 1:6){
+ lines(x = rep(i,2),
+ lwd = 3,
+ col = "#666666",
+ y = c(max(ss.data.pc$pc.volume[ss.data.pc$pc.batch==i]),
+ min(ss.data.pc$pc.volume[ss.data.pc$pc.batch==i])))}
> abline(h = 16,
+ lwd = 2)
```



## 8 Diagrama de Caja

Resume gráficamente la distribución de una variable continua. Los lados de la caja son la primera y tercera cuartiles (25vo y 75vo percentiles, respectivamente). Por lo tanto, dentro de la caja que tenemos el medio 50% de los datos. La mediana se representa como una línea que atraviesa la caja. Los valores extremos pueden ser los máximos y mínimos de los datos u otros límites más allá de los cuales los datos se consideran valores atípicos. El diagrama de caja nos dice si la distribución está centrada o sesgada (la posición de la mediana con respecto al resto de los datos), si hay valores atípicos (puntos fuera de los extremos), o si los datos son cercanos a los valores centrales.

```
> my.bp <- boxplot(ss.data.pc$pc.volume,
+ col = "#CCCCCC",
+ main = "Box Plot of Volume",
+ sub = "Printer Cartridge Example",
+ ylab = "Volume",
+ range = 0.7)
> text(x = rep(1, length(my.bp$out)),
+ y = my.bp$out,
+ labels = which(ss.data.pc$pc.volume==my.bp$out),
```

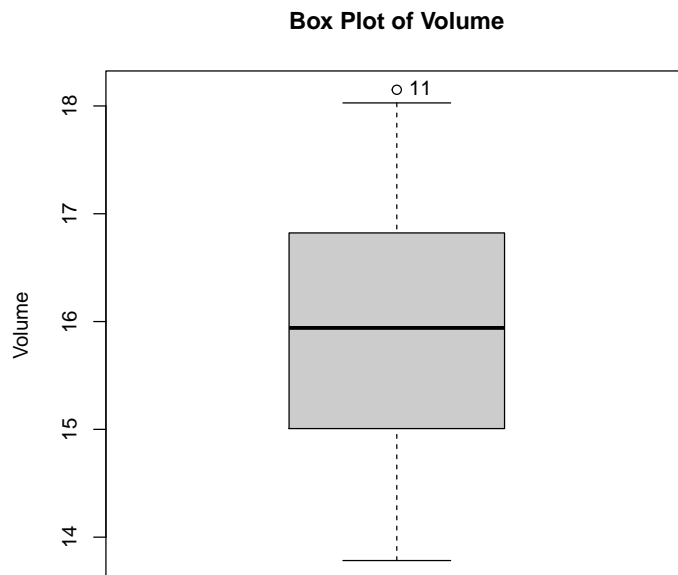
```

+ pos = 4)
> str(my.bp)

List of 6
 $ stats: num [1:5, 1] 13.8 15 15.9 16.8 18
 $ n      : num 24
 $ conf   : num [1:2, 1] 15.4 16.5
 $ out    : num 18.2
 $ group  : num 1
 $ names  : chr ""

>

```



Printer Cartridge Example

Ejemplo 7. Vamos a trazar un diagrama de caja para todos los datos con el siguiente código, considerando para los valores atípicos un rango de 0.7. Se obtiene un valor atípico, correspondiente a la fila 11 del conjunto de datos. Para el ejemplo, utilizamos la función `boxplot`.

## 9 Gráfica de Grupos

En un gráfico de grupo, identificamos los puntos en el gráfico, mostrando el grupo al que pertenecen. Es útil cuando la característica que estamos midiendo es producida por diferentes corrientes de proceso o se mide en diferentes lugares.

Los valores más altos y más bajos de cada grupo están generalmente unidos por líneas. **Ejemplo 8.** Supongamos que los cuatro cartuchos de toner de impresora en cada lote son etiquetados de acuerdo con el operador que vende el cartucho (A, B, C, y D). La información está contenida en la variable de `pc.op` del paquete SixSigma. Con el siguiente código obtenemos el gráfico de grupo en la fig. 8. Utilizamos la función `stripchart`.

```
> stripchart(pc.volume ~ pc.batch,
+ vertical = TRUE,
+ data = ss.data.pc,
+ pch="",
+ xlab = "Batch (lote)",
+ ylab = "Volumen",
+ ylim = c(12,20),
+ main = "Agrupado por (Operator) Carta de Volumen",
+ sub = "Printer Cartridge Example")
> rect(par("usr")[1],
+ par("usr")[3],
+ par("usr")[2],
+ par("usr")[4],
+ col = "#CCCCCC")
> box(col = "#CCCCCC")
> grid(col = "#EEEEEE")
> points(pc.volume ~ pc.batch, data = ss.data.pc, pch = 19)
> with(ss.data.pc, text(label = pc.op,
+ x = pc.batch,
+ y = pc.volume, pos = 4))
> lines(aggregate(pc.volume ~ pc.batch, data = ss.data.pc, max), lwd=2)
> lines(aggregate(pc.volume ~ pc.batch, data = ss.data.pc, min),
+ lwd=2)
>
```

Agrupado por (Operator) Carta de Volumen

