

# Metodología Six Sigma

Herramientas para la fase de análisis: Gráficos con R

*Profesor: Gustavo Ahumada*

## Introducción

Los gráficos son particularmente importantes en los proyectos Six Sigma. El objetivo de un gráfico suele ser apoyar la interpretación de los datos. Por lo tanto, proporcionar una explicación adecuada de los datos a través de gráficos es crucial. Los gráficos bidimensionales se pueden utilizar como base para representaciones más complejas, y los datos multidimensionales se pueden mostrar a través de extensiones de estos gráficos.

**Ejemplo** Cartuchos de impresora. La base de datos `ss.data.pc` viene por defecto en el paquete **SixSigma**.

```
library(SixSigma)
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
## 7	C	1	14.69381	1.345546	2	C
## 8	C	2	15.19748	1.184424	2	D
## 9	C	3	14.20776	1.457567	3	A
## 10	C	1	15.95797	1.186025	3	B
## 11	C	2	18.15004	1.276415	3	C
## 12	C	3	14.22977	1.280528	3	D
## 13	B	1	16.86465	1.301196	4	A
## 14	B	2	14.27984	1.112285	4	B
## 15	B	3	16.13413	1.270603	4	C
## 16	B	1	15.92417	1.341316	4	D
## 17	B	2	16.85830	1.199093	5	A
## 18	B	3	16.34490	1.199544	5	B
## 19	B	1	15.41755	1.541008	5	C
## 20	B	2	16.78617	1.312416	5	D
## 21	B	3	15.30790	1.142107	6	A
## 22	B	1	14.81696	1.216941	6	B
## 23	B	2	17.26885	1.260321	6	C
## 24	B	3	15.68745	1.329866	6	D

```
plot(pc.volume ~ pc.density,
data = ss.data.pc,
pch = 16)
```

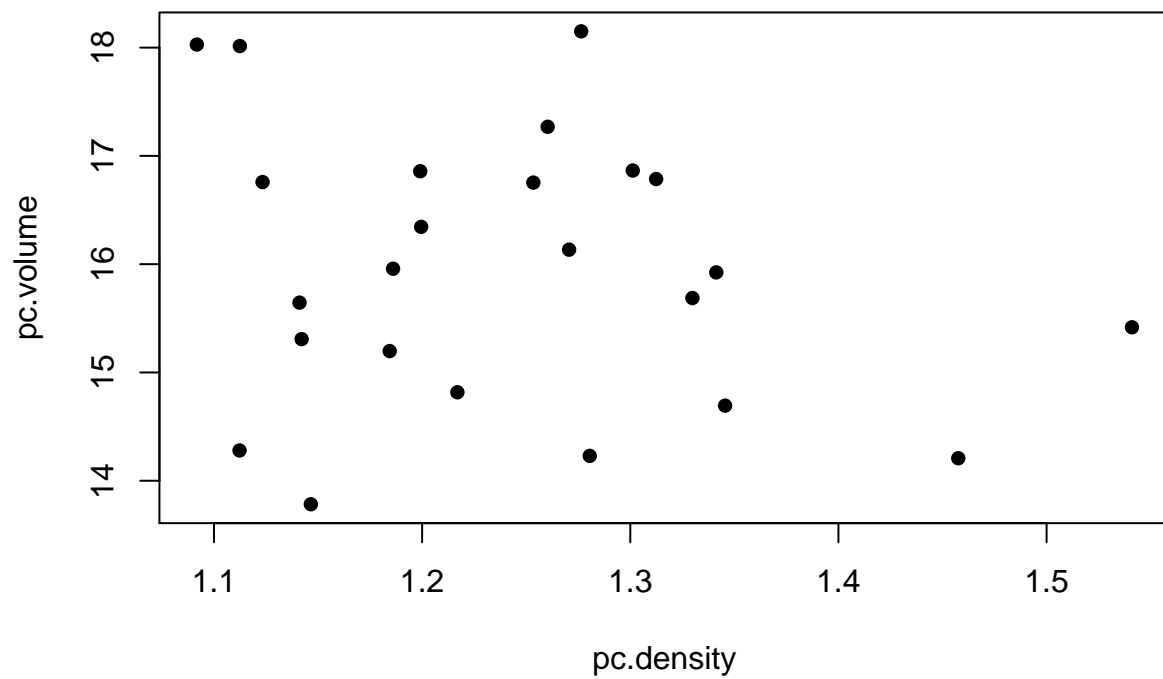
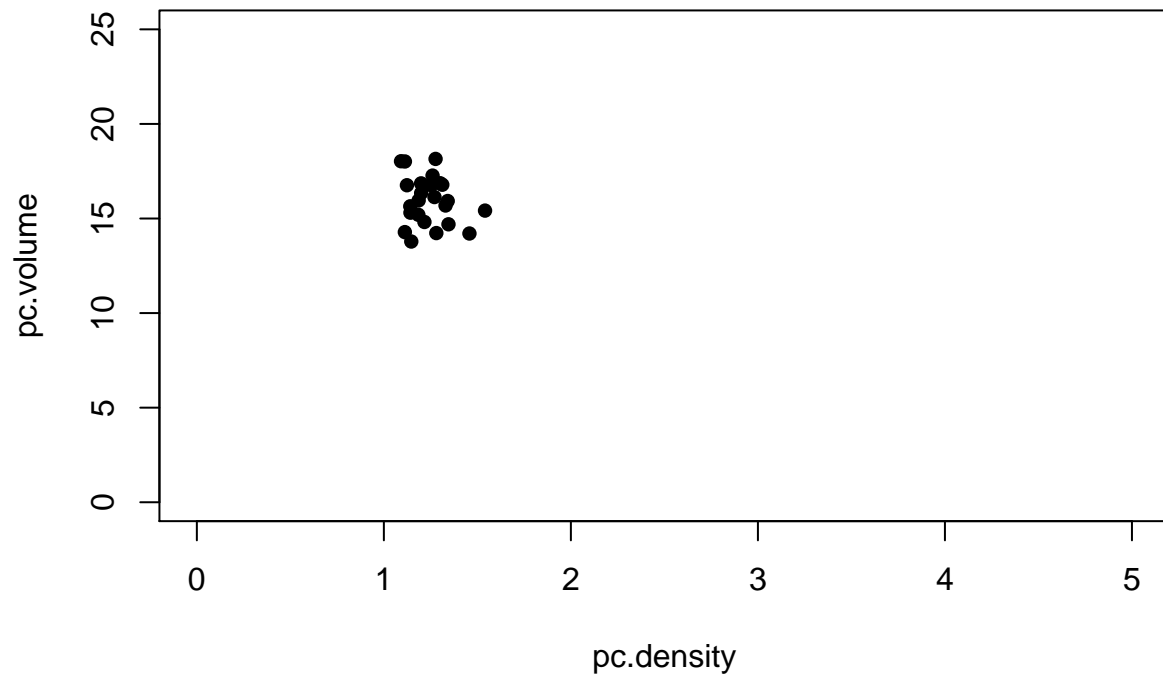


Gráfico con los límites por defecto cambiados

```
plot(pc.volume ~ pc.density,  
data = ss.data.pc,  
xlim = c(0,5),  
ylim = c(0,25), pch=16)
```



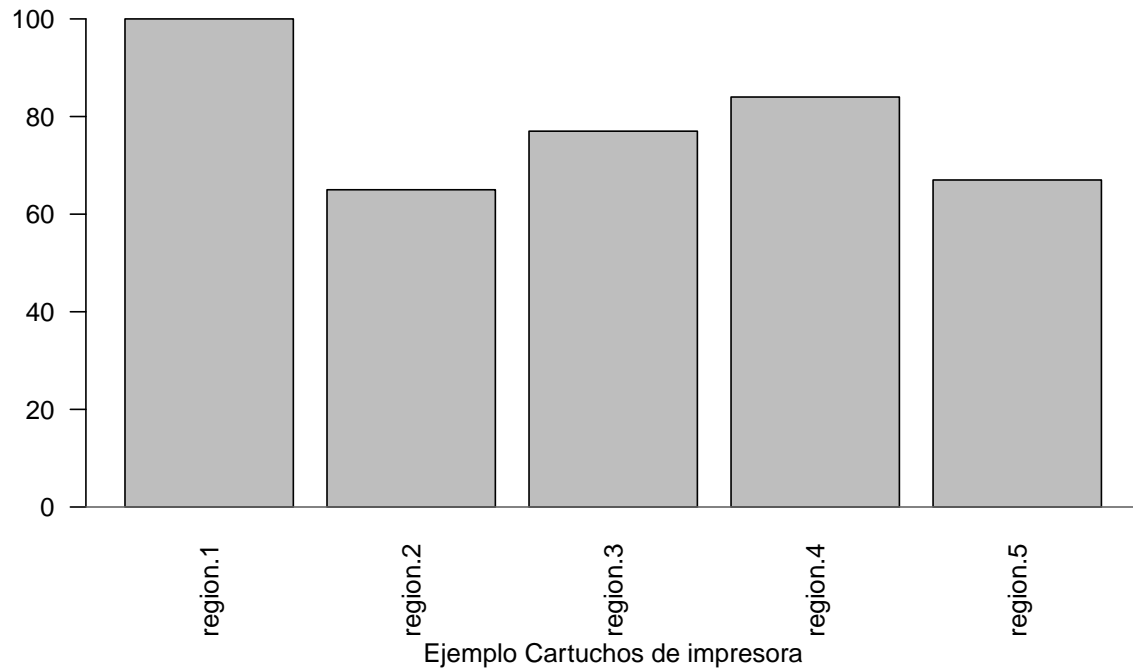
## Gráfico de barras

Un gráfico de barras es un gráfico muy simple donde algunas cantidades se muestran como la altura de las barras. Cada barra representa un factor donde se mide la variable en estudio. Un gráfico de barras suele ser la mejor representación gráfica para los recuentos.

**Ejemplo** Cartuchos de impresora (continuación). El productor de cartuchos de impresora reparte su producto en cinco regiones. Una cantidad inesperada de cartuchos defectuosos han sido retornados en el último mes, a continuación tenemos un gráfico de barras del número de cartuchos defectuosos por región:

```
with(ss.data.pc.r,
barplot(pc.def,
names.arg = pc.regions,
las = 2,
main = "Gráfico de barra de cartuchos defecuosos por región",
sub = "Ejemplo Cartuchos de impresora"))
abline(h = 0,
col = "#666666")
```

**Gráfico de barra de cartuchos defecuosos por región**



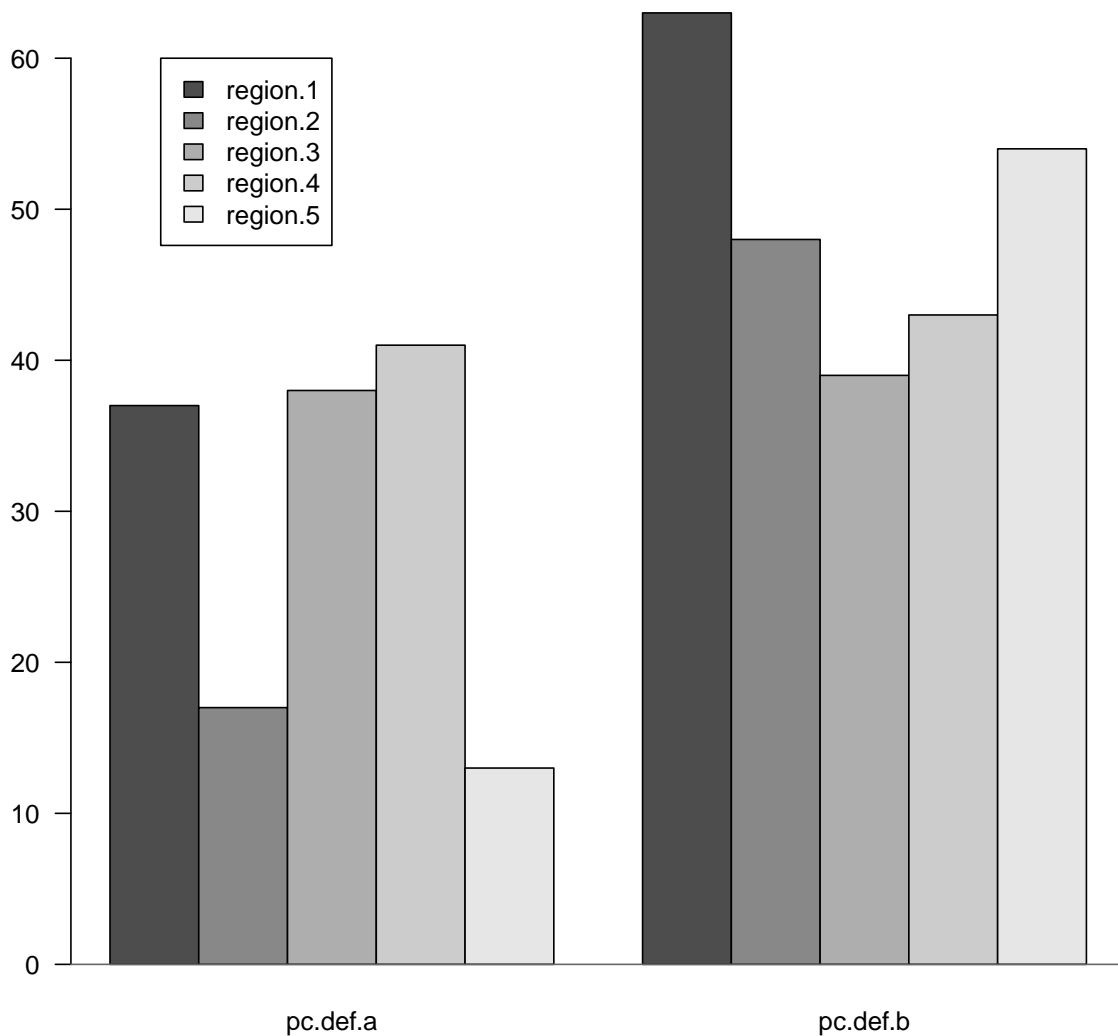
```
ss.data.pc.r
```

```
##   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
```

**Ejemplo** Cartuchos de impresora (continuación). Se puede graficar un gráfico de barras para cada tipo de cartucho en el mismo gráfico, aún con un leyenda, empleando el siguiente comando:

```
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 barra de cartuchos defecuosos por región y tipo",
sub = "Ejemplo Cartuchos de impresora")
abline(h = 0,
col = "#666666")
```

**Gráfico de barra de cartuchos defecuosos por región y tipo**



**Ejemplo Cartuchos de impresora**

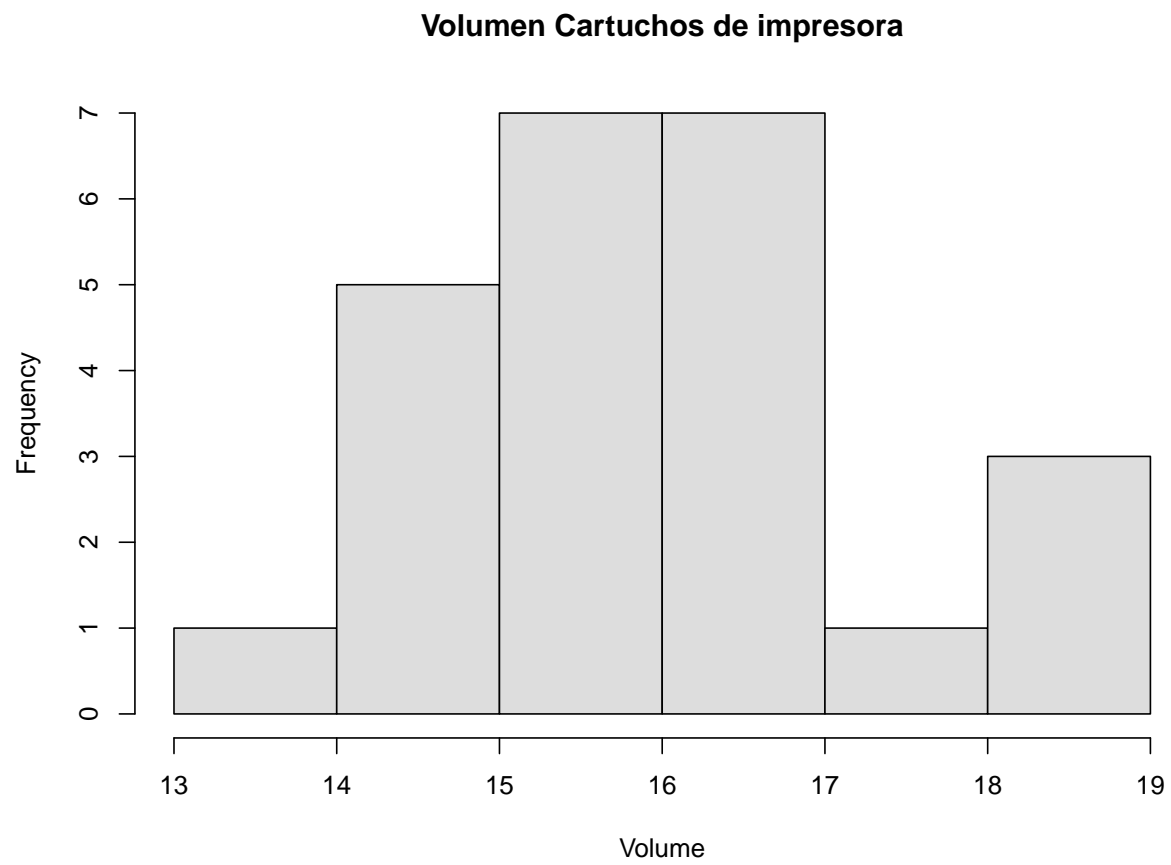
## Histograma

Un histograma es un gráfico de barras para variables continuas. Este gráfico de barras muestra la distribución de las medidas de variables. En el eje x, cada barra representa un intervalo de los posibles valores 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.

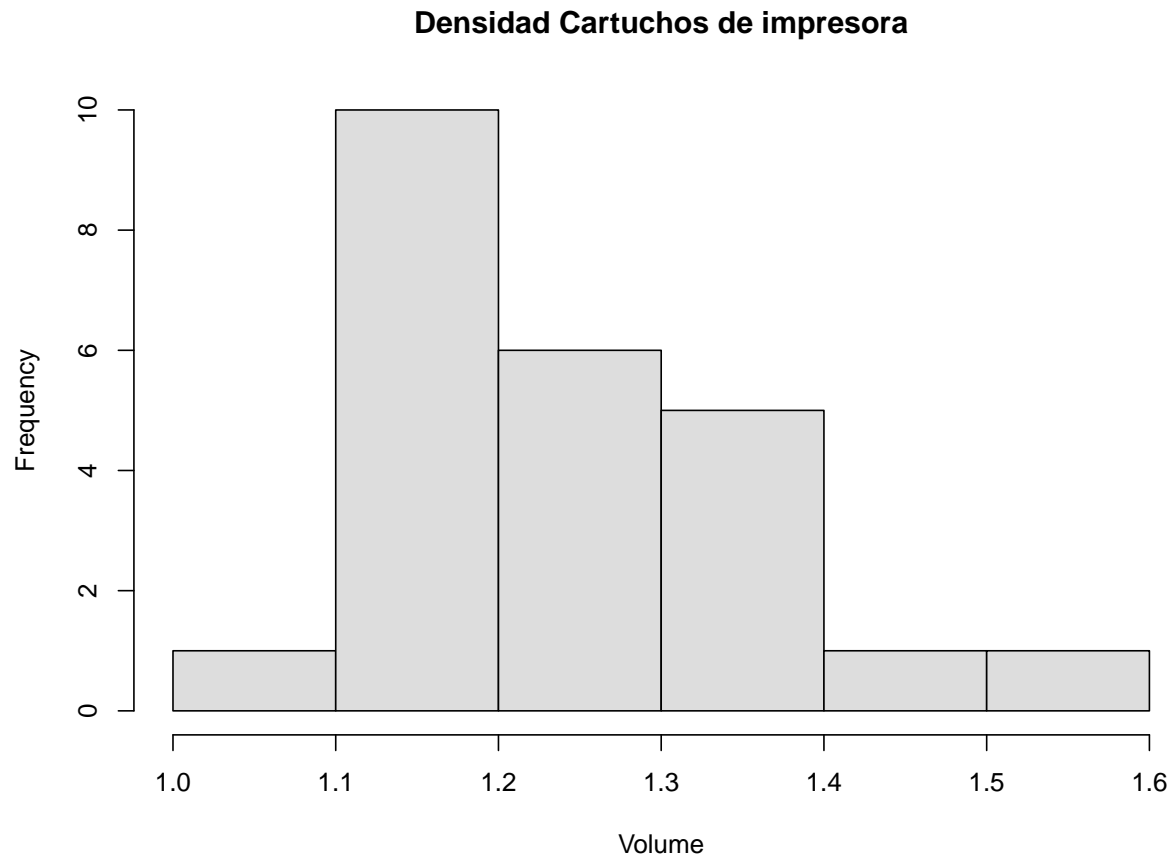
**Ejemplo** Cartuchos de impresora (continuación). La siguiente figura tiene un histograma para las variables *volume* y *density* en la base de datos **ss.data.pc**. El siguiente código es utilizada para ello:

```
hist(ss.data.pc$pc.volume,  
main="Volumen Cartuchos de impresora",  
xlab="Volume",
```

```
col="#DDDDDD")
```



```
hist(ss.data.pc$pc.density,  
breaks = "FD",  
main = "Densidad Cartuchos de impresora",  
xlab = "Volume",  
col = "#DDDDDD")
```



## Diagrama de dispersión

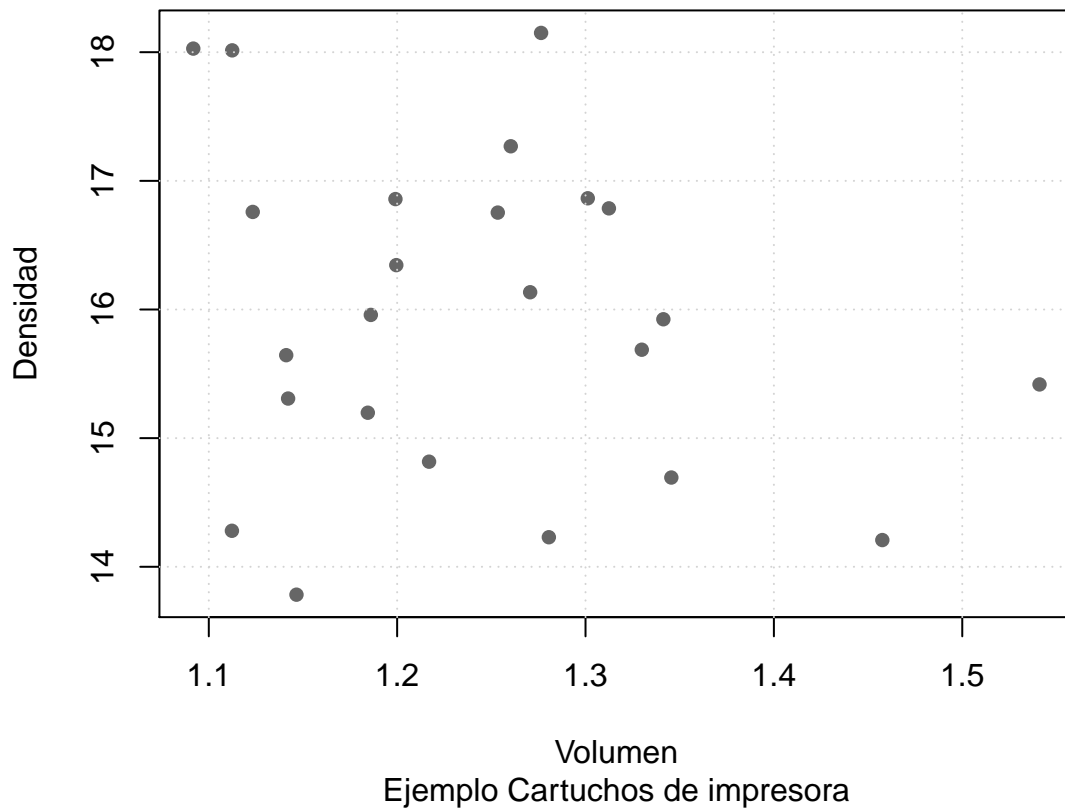
Un diagrama de dispersión es una herramienta importante para revelar relaciones entre dos variables. En lenguaje estadístico, estas relaciones se pueden incluir en el concepto de correlación. Por lo tanto, podemos tener tres tipos de correlación entre dos variables

Entre un proyecto Six Sigma, descubrir la relación entre las  $Y$ s y las  $X$ s del proceso es una tarea importante. Esta relación será medida en términos de la correlación entre las variables. Una descubierta la relación, se deberá probar que la variable independiente está causando la variación de la variable dependiente.

**Ejemplo** Cartuchos de impresora (continuación). En el conjunto de datos **ss.data.pc**, se tiene dos variables continuas **pc.volume** y **pc.density**. Si se quiere verificar que la densidad y el volumen están correlacionadas, la primera cosa que se debe hacer es generar

```
plot(pc.volume ~ pc.density,
data = ss.data.pc,
main = "Buscando correlación entre Densidad y Volumen",
col = "#666666",
pch = 16,
sub = "Ejemplo Cartuchos de impresora",
xlab = "Volumen",
ylab = "Densidad")
grid()
```

## Buscando correlación entre Densidad y Volumen



## Gráficos de series de tiempo

El gráfico de serie de tiempo es un gráfico bidimensional donde el eje  $x$  representa una línea de tiempo y en el eje  $y$  se traza una variable que queremos monitorear. Esta variable puede ser una característica crítica para la calidad (CTQ) de nuestro proceso o un parámetro que lo afecte. Estos tipos de gráficos también se denominan gráficos de series de tiempo cuando tenemos una escala de tiempo en el eje  $x$  (por ejemplo, la cantidad de pedidos recibidos cada día). La escala del eje  $x$  puede no ser necesariamente temporal (por ejemplo, el volumen de algunos destinatarios cuya producción es secuencial)

**Ejemplo** Cartuchos de impresora (continuación). Se quiere monitorear el volumen en los cartuchos, disponibles en `ss.data.pc`. Suponer que el valor objetivo para el volumen es 16.

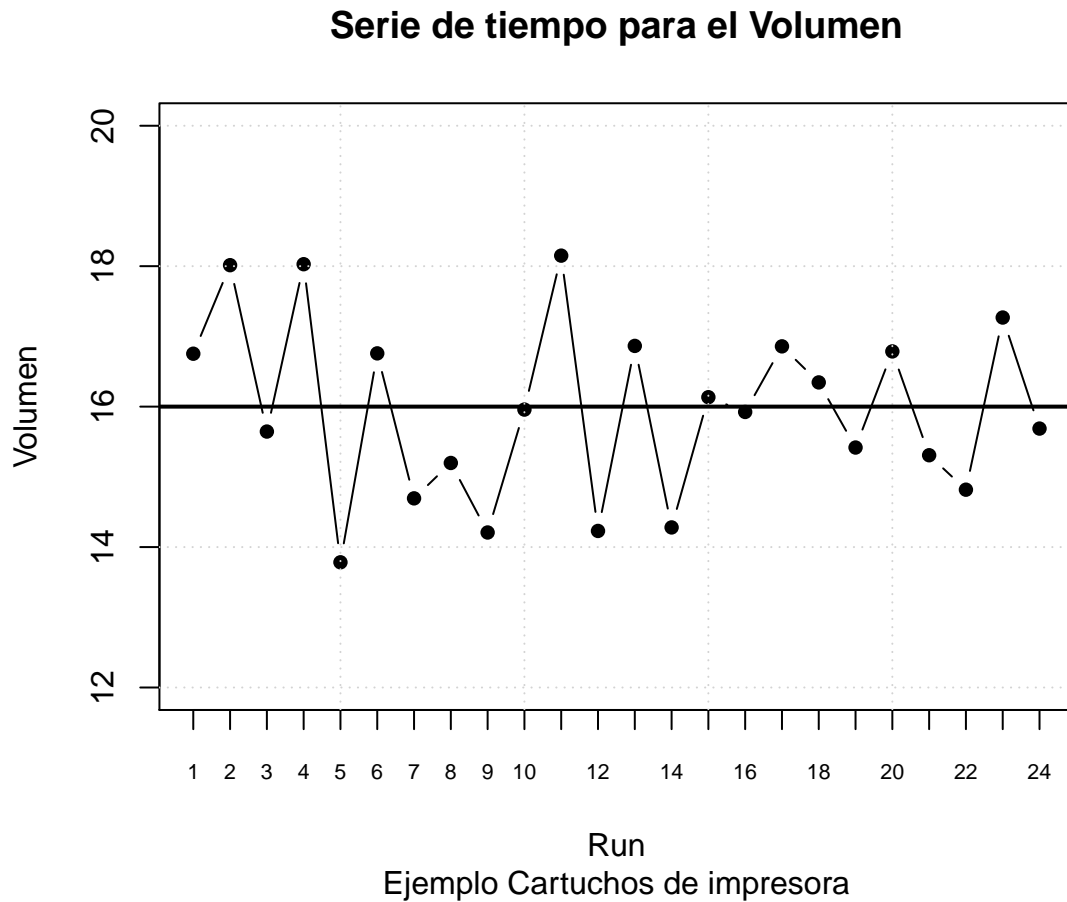
```
plot(ss.data.pc$pc.volume,
type = "b",
pch = 16,
ylim = c(12,20),
axes = FALSE,
main = "Serie de tiempo para el Volumen",
sub = "Ejemplo Cartuchos de impresora",
xlab = "Run",
ylab = "Volumen")
axis(1,
```



```

at = 1:24,
cex.axis = 0.7)
axis(2)
box()
grid()
abline(h = 16,
lwd = 2)

```



## Gráfico de Caja y bigotes

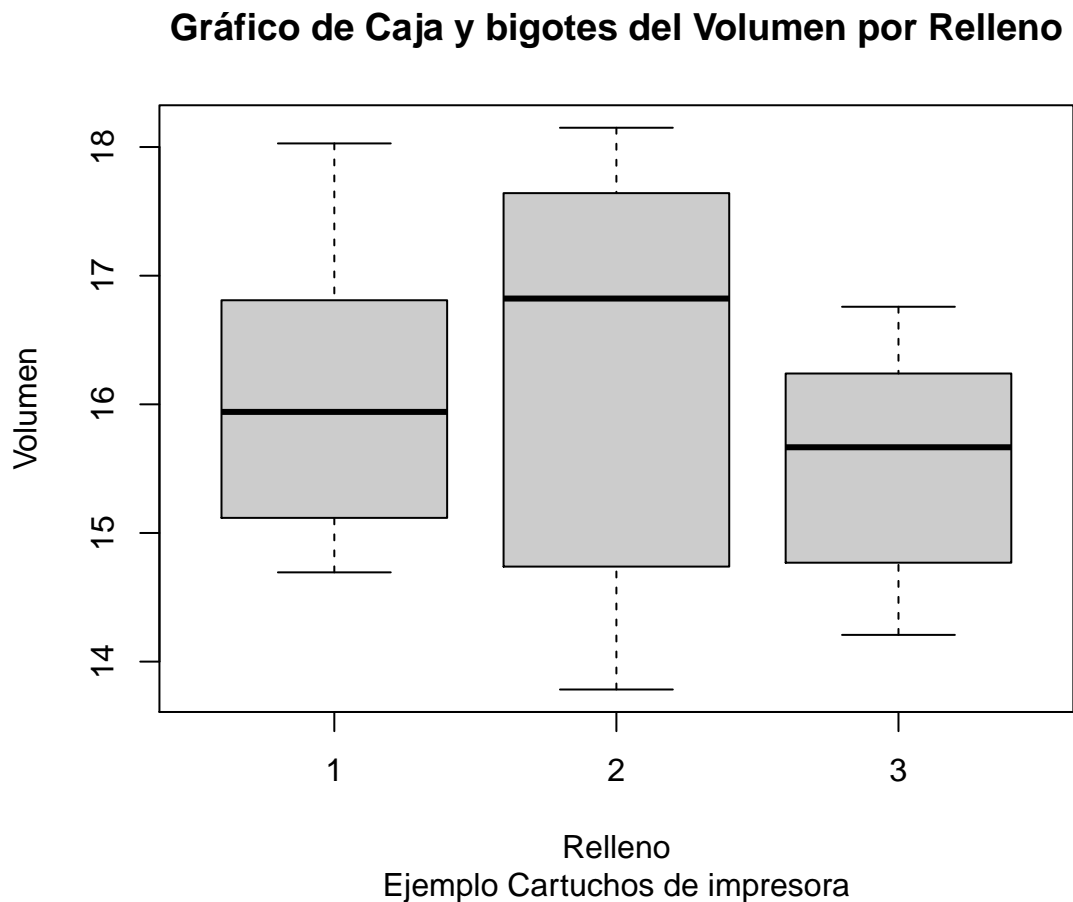
El gráfico de caja y bigotes también se conoce como el diagrama de caja. Resume gráficamente la distribución de una variable continua. Los lados de la caja son los cuartiles primero y tercero (percentil 25 y 75, respectivamente). Por lo tanto, dentro del cuadro tenemos el 50% de los datos. La mediana se traza como una línea que cruza el cuadro. Los valores extremos de los bigotes pueden ser el máximo y el mínimo de los datos u otros límites más allá de los cuales los datos se consideran atípicos. Los límites generalmente se toman como:

$$[Q1 - 1.5 \times IQR, Q3 + 1.5 \times IQR],$$

donde  $Q1$  y  $Q3$  son el primer y el tercer cuartil, respectivamente, y  $IQR$  es el rango intercuartil ( $Q3 - Q1$ )

**Ejemplo** Cartuchos de impresora (continuación). En una línea de producción, tenemos tres rellenos para los cartuchos de impresora. Queremos determinar si hay alguna diferencia entre los rellenos e identificar valores atípicos (por ejemplo, errores en la entrada de datos). Obtenemos el diagrama de caja para cada grupo con el siguiente código

```
boxplot(pc.volume ~ pc.filler,  
data = ss.data.pc,  
col = "#CCCCCC",  
main = "Gráfico de Caja y bigotes del Volumen por Relleno",  
sub = "Ejemplo Cartuchos de impresora",  
xlab = "Relleno",  
ylab = "Volumen")
```



Entro otras más