

Datos Anómalos

Paúl Arévalo

2024-07-04

Datos Atípicos

Ejemplo

A continuación se genera un conjunto de datos con valores atípicos

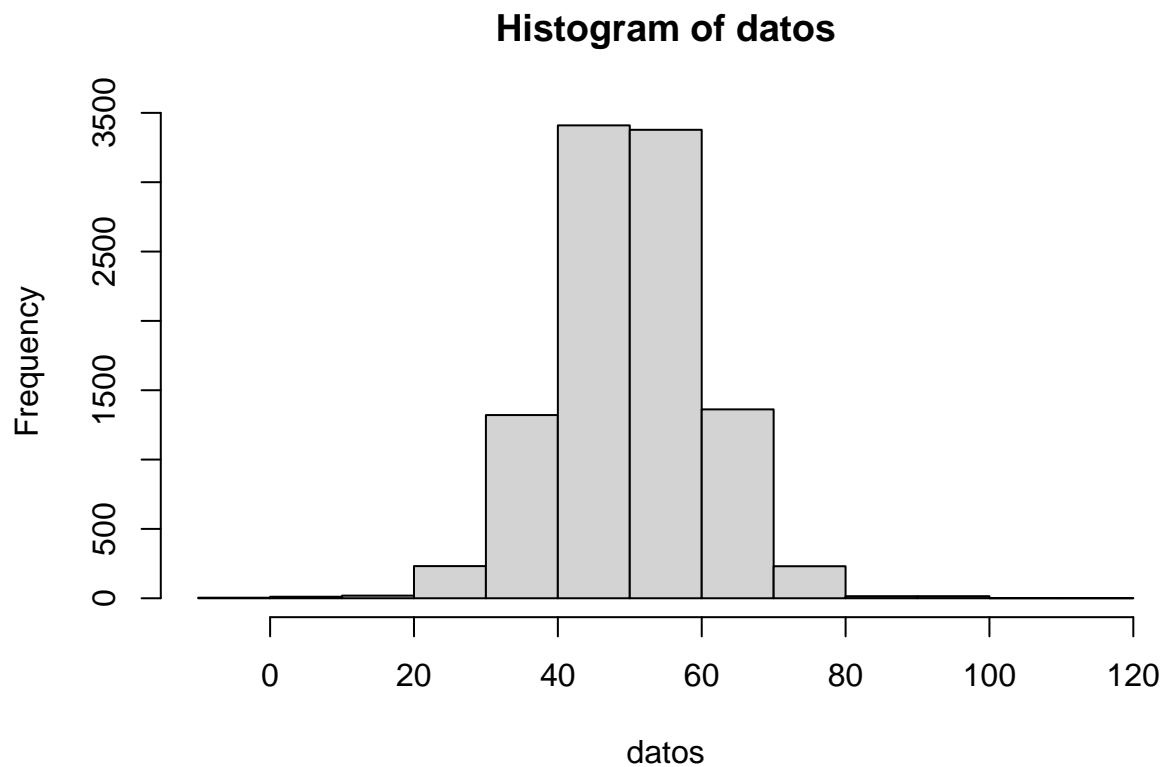
```
set.seed(12345)
# Generar distribución normal con media 50 y desviación estándar 10
datos <- rnorm(10000, mean = 50, sd = 10)
# Introducir valores atípicos
for (i in 1:100) {
  index <- sample.int(length(datos), 1)
  datos[index] <- datos[index] + runif(n = 1, min = -50, max = 50)
}
```

Realizamos una análisis exploratorio de datos:

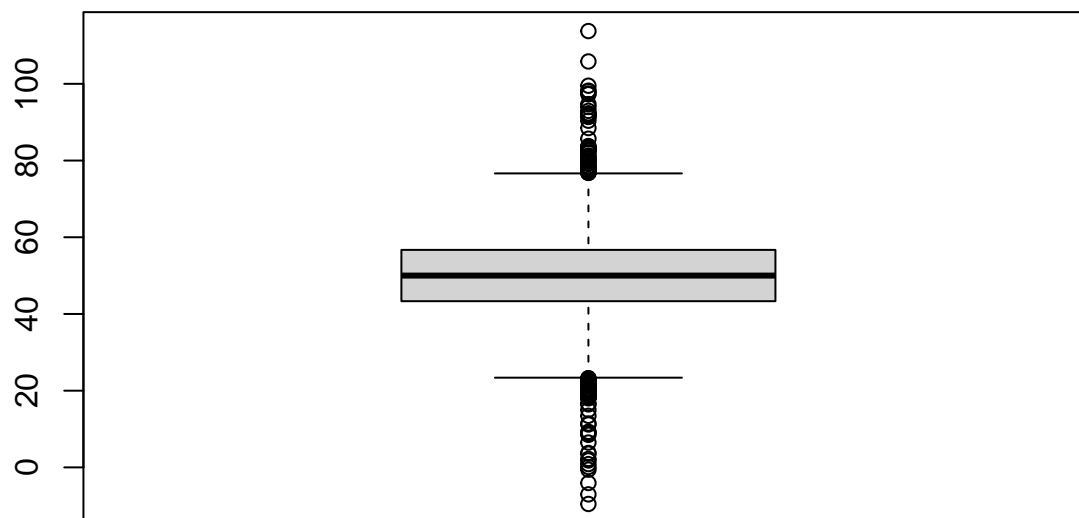
```
# Observar la distribución de los datos
summary(datos)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -9.535  43.330   50.005   50.021  56.704  113.757
```

```
hist(datos)
```

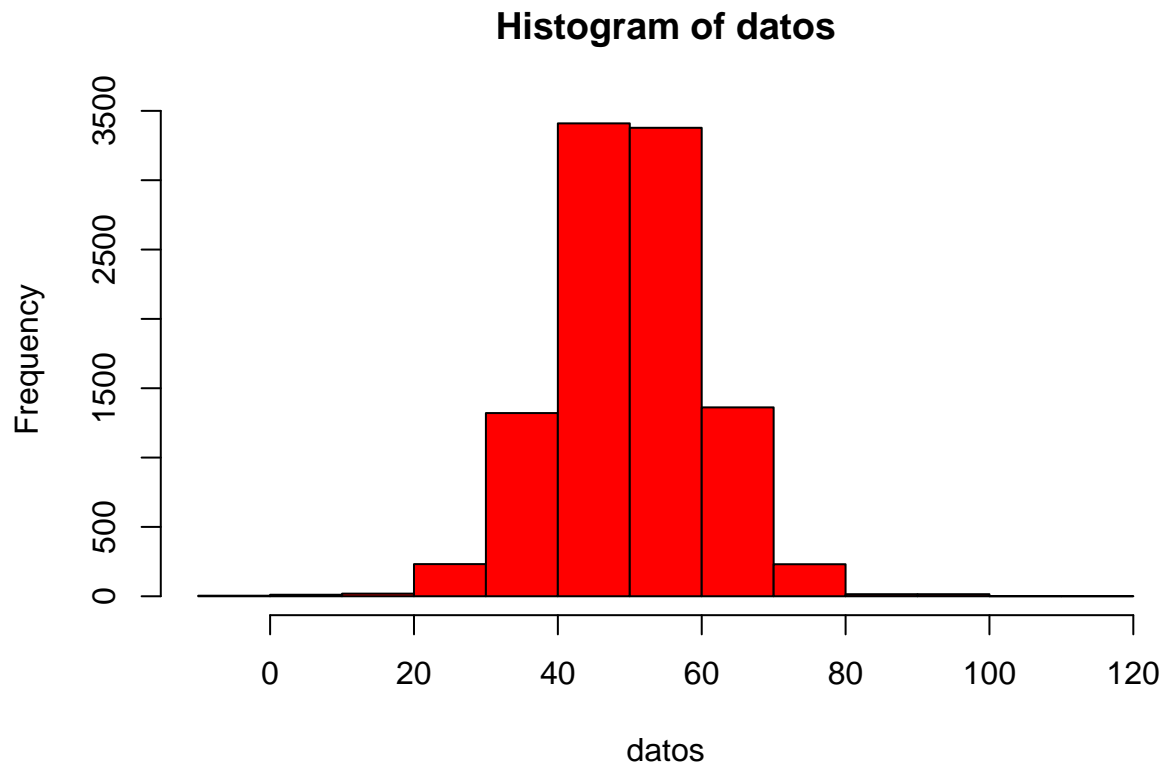


```
boxplot(datos)
```



Procedemos a detectar valores atípicos

```
# Utilizar el método IQR (Rango intercuartílico)
iqr <- IQR(datos)
limite_superior <- median(datos) + 1.5 * iqr
limite_inferior <- median(datos) - 1.5 * iqr
outliers <- datos[datos > limite_superior | datos < limite_inferior]
# Identificar outliers en el histograma
hist(datos, col = ifelse(is.na(outliers), "lightblue", "red"))
```



```
plot(datos, main = "Diagrama de dispersión con outliers")  
points(outliers, col = "red", pch = 16, cex = 2)
```

