

Clasificación de Salmón del Atlantico con destino a filete

Diplomado de Análisis de datos con R para la Acuicultura

Cristian Naguian Asenjo

10 May 2022

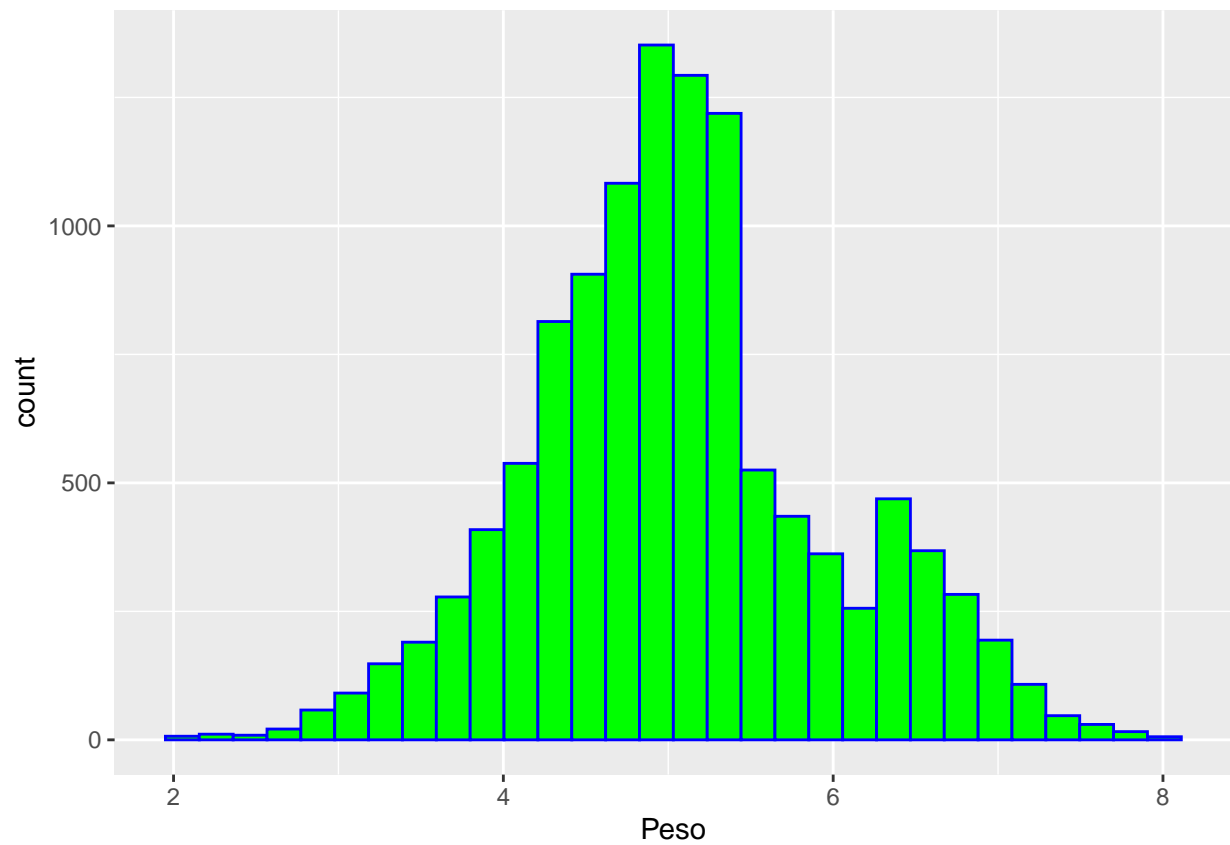
Medidas de resumen).

```
set.seed(1) #semilla para fijar resultados cada vez que se corre la simulación
datos <- read.delim("/cloud/project/Piezas a filete.txt", na="NA")
str(datos)

## 'data.frame':   11526 obs. of  5 variables:
## $ Pieza   : int  1 2 3 4 5 6 7 8 9 10 ...
## $ Peso    : num  3.45 3.78 3.79 3.9 3.69 ...
## $ Largo   : num  0.635 0.635 0.635 0.61 0.635 ...
## $ Calibre: chr   "2.7-4.0 Kg" "2.7-4.0 Kg" "2.7-4.0 Kg" "2.7-4.0 Kg" ...
## $ Calidad: chr   "Premium" "Premium" "Premium" "Premium" ...

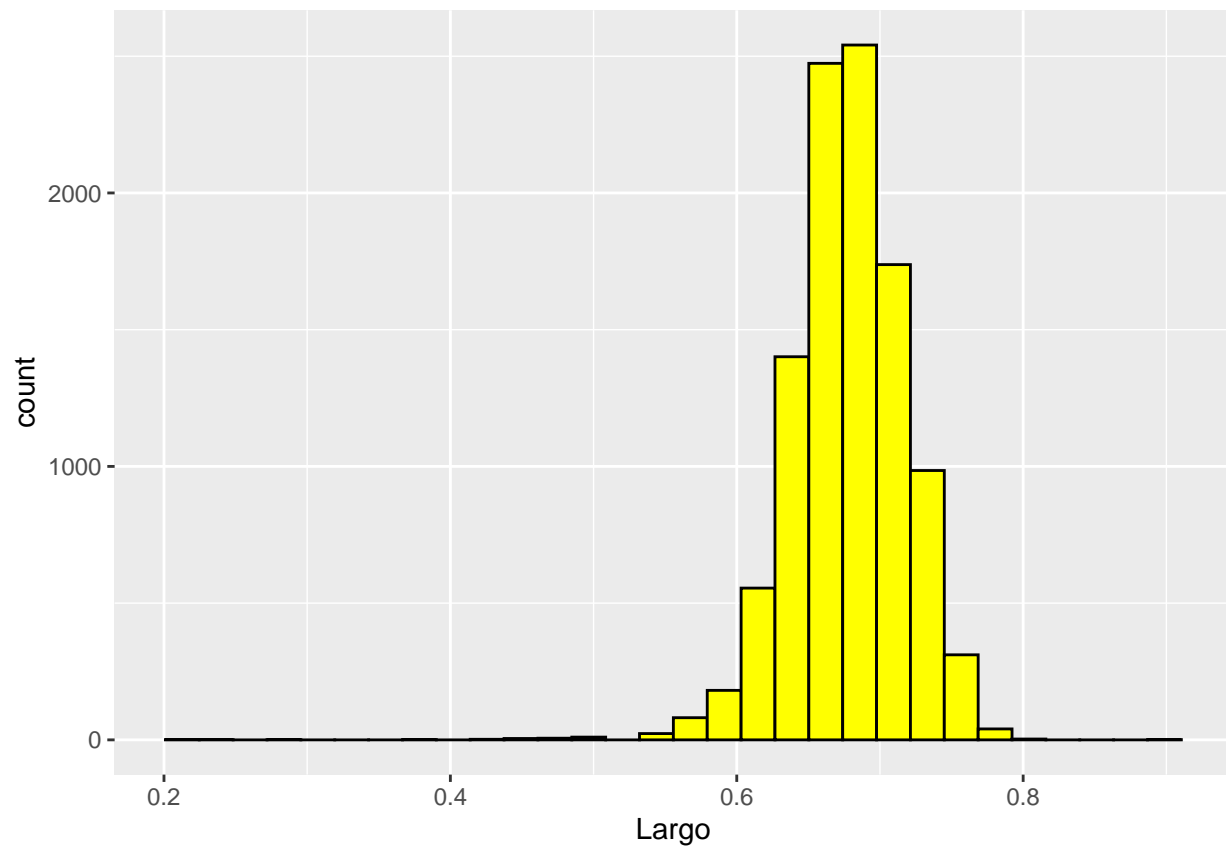
## Describe la variación de las variables de estudio usando histogramas

```r
ggplot(datos, aes(x = Peso)) +geom_histogram(bins = 30, color = "blue", fill="green")
```



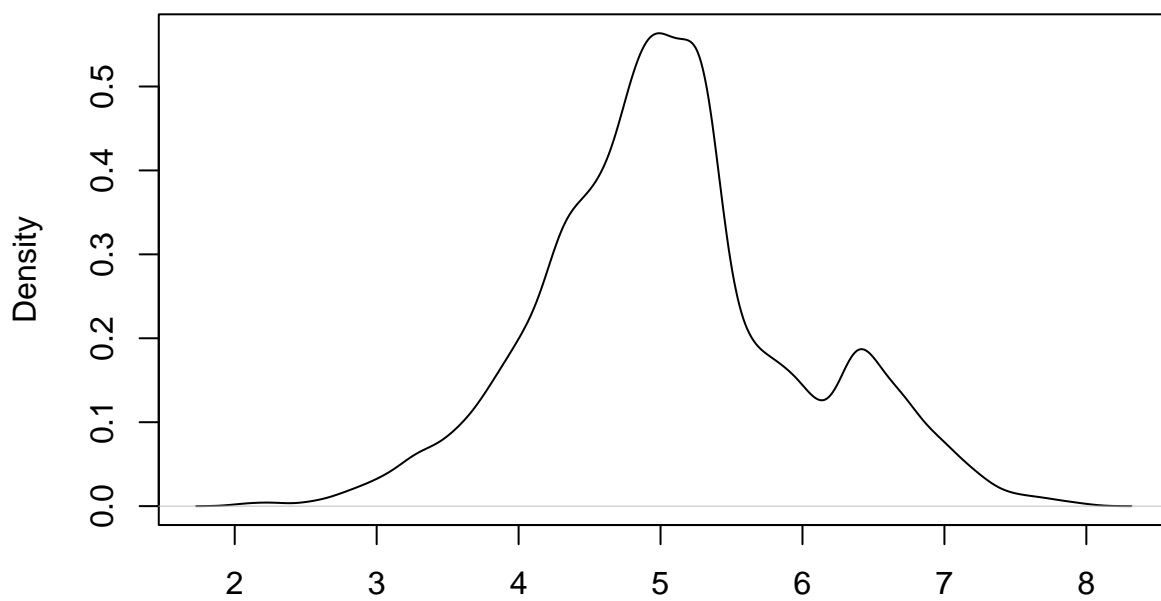
```
ggplot(datos, aes(x = Largo)) +geom_histogram(bins = 30, color = "black", fill="yellow")
```

```
Warning: Removed 1165 rows containing non-finite values (stat_bin).
```



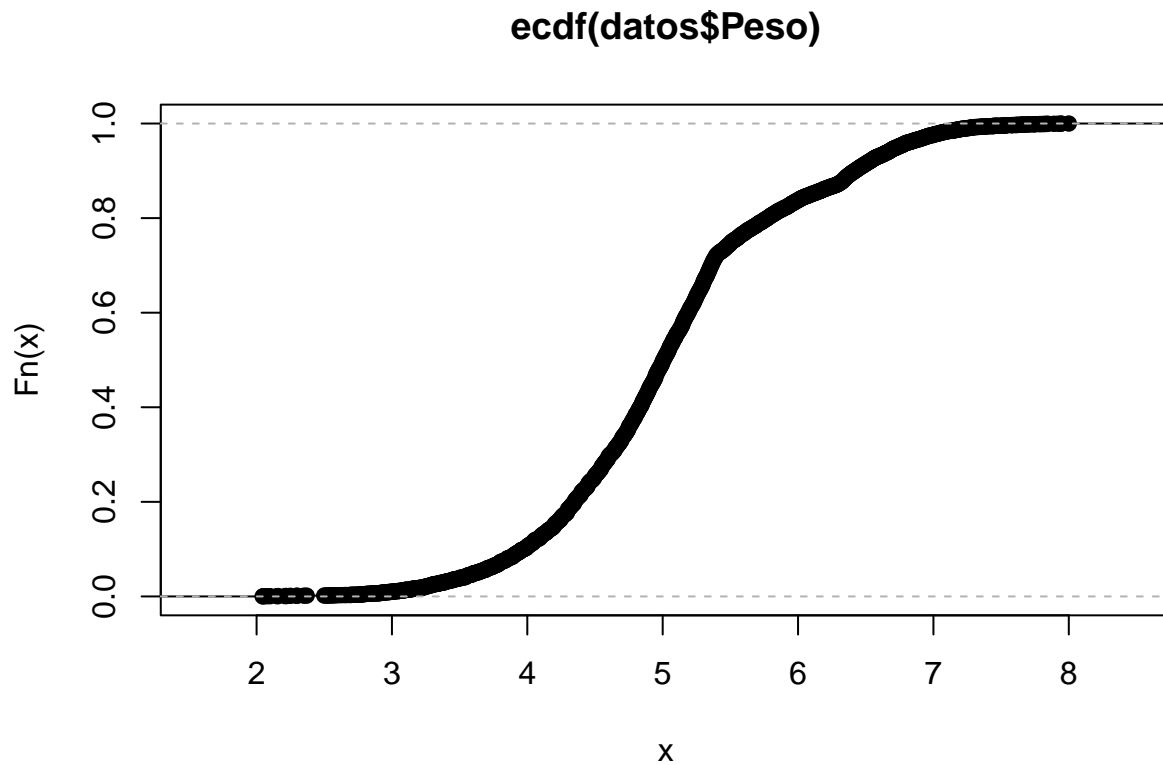
```
plot(density(datos$Peso))
```

**density.default(x = datos\$Peso)**



N = 11526 Bandwidth = 0.1055

```
plot(ecdf(datos$Peso))
```



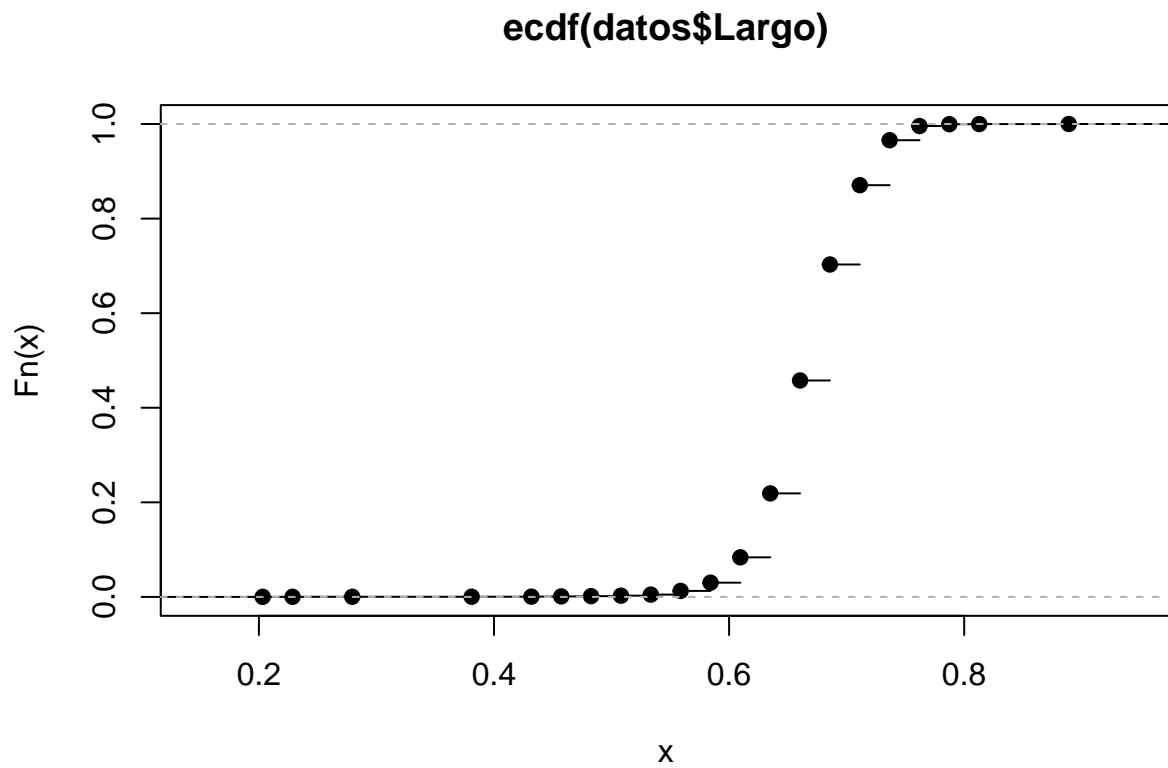
```
ecdf(datos$Peso) #Distribución empírica acumulada de la variable weight.
```

```
Empirical CDF
```

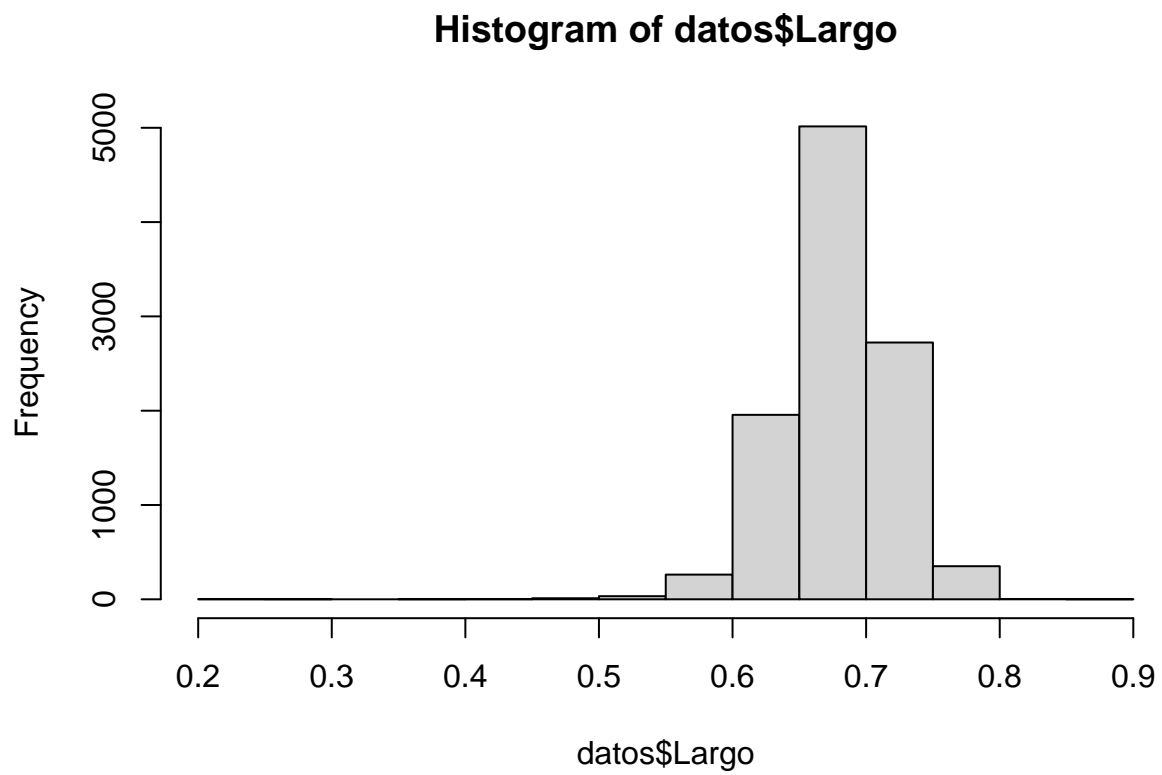
```
Call: ecdf(datos$Peso)
```

```
x[1:982] = 2.045, 2.05, 2.075, ..., 7.945, 8
```

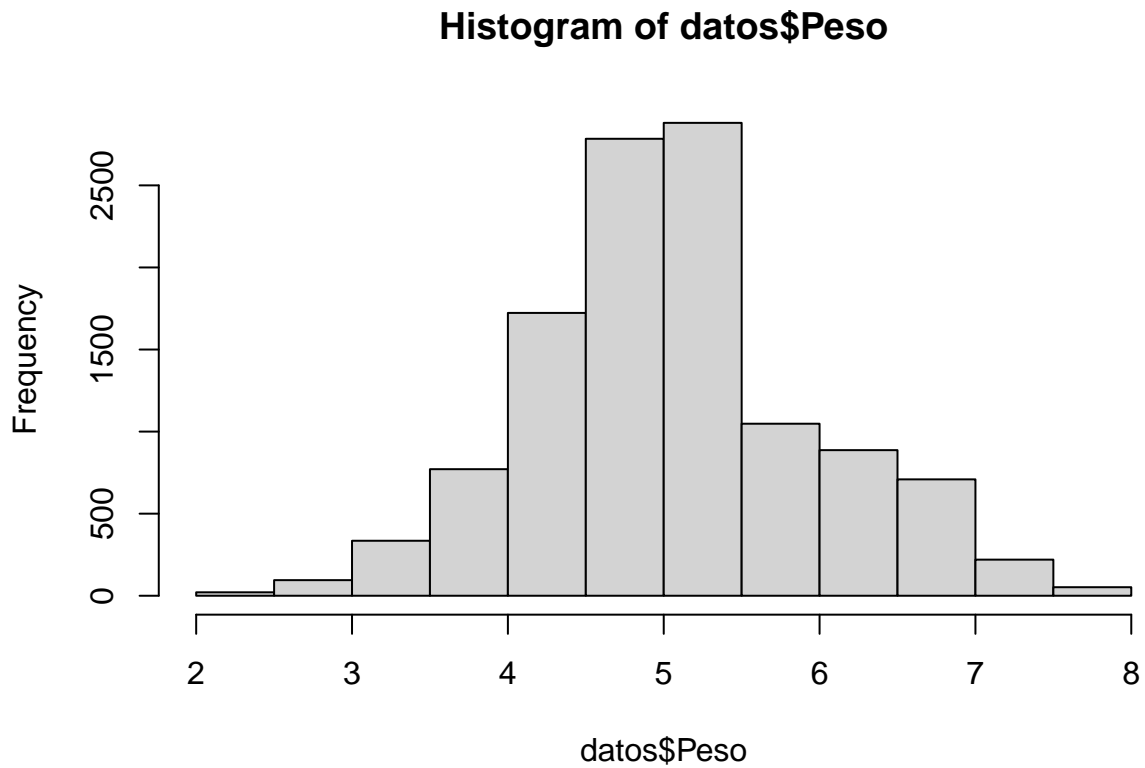
```
plot(ecdf(datos$Largo))
```



```
hist(datos$Largo)
```

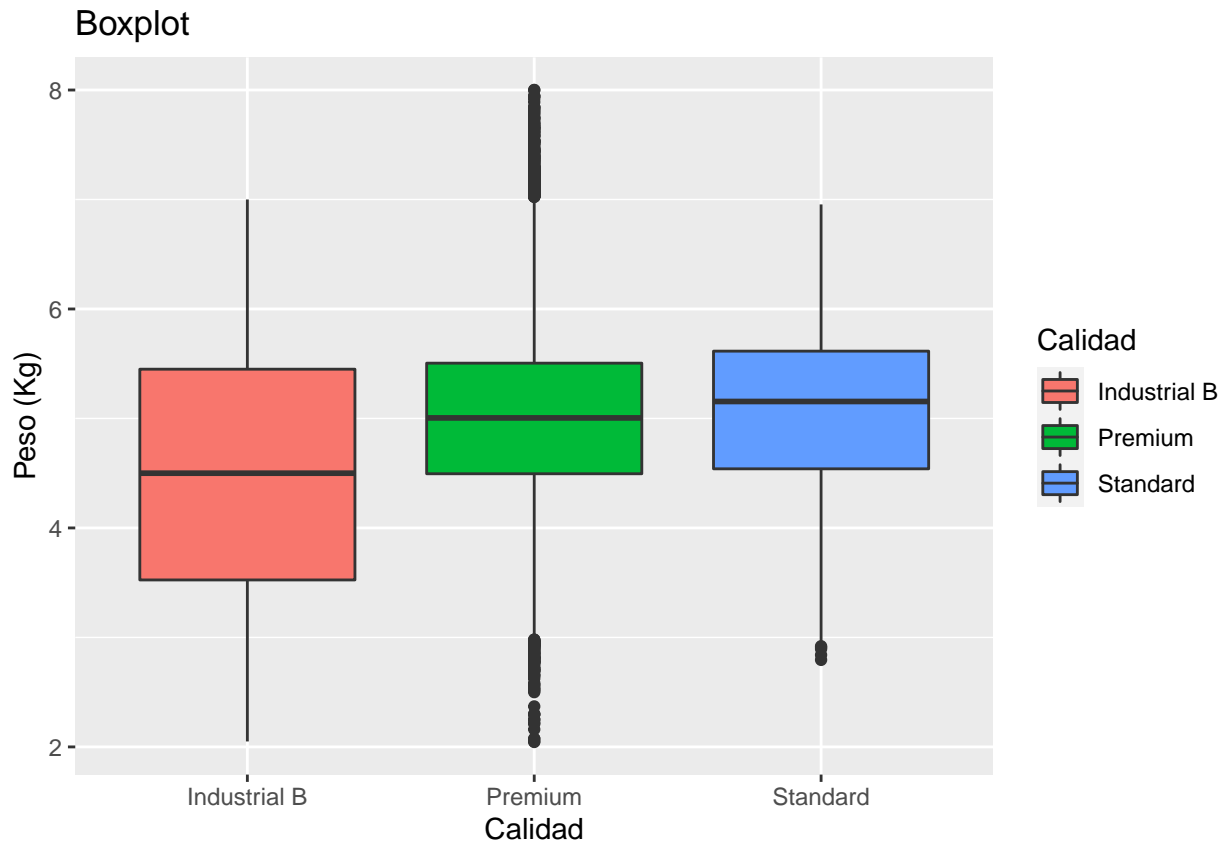


```
hist(datos$Peso)
```



Establece relación entre variables cuantitativas y factores usando gráficas de correlación, boxplot, interacción o de tamaño de los efectos

```
ggplot(datos, aes(x=Calidad, y=Peso, fill = Calidad)) +geom_boxplot()+labs(title="Boxplot", x="Calidad"
```



## Identifica si existen errores, datos faltantes o valores atípicos

```
summary(datos)
```

```
Pieza Peso Largo Calibre
Min. : 1 Min. :2.045 Min. :0.2032 Length:11526
1st Qu.: 2882 1st Qu.:4.495 1st Qu.:0.6604 Class :character
Median : 5764 Median :5.010 Median :0.6858 Mode :character
Mean : 5764 Mean :5.069 Mean :0.6769
3rd Qu.: 8645 3rd Qu.:5.515 3rd Qu.:0.7112
Max. :11526 Max. :8.000 Max. :0.8890
NA's :1165
##
Calidad
Length:11526
Class :character
Mode :character
##
##
##
##
```

Como se puede observar en el resumen de datos, existen 1156 datos de largo los cuales no estan ingresados, debido a que la grader, maquina que calibra las piezas no pude detectar su longitud.

## Resume los datos usando tablas y estadística descriptiva

```
table(datos$Calidad)
```

```

Industrial B Premium Standard
71 10898 557
```

```
table(datos$Calibre)
```

```

2-3 Kg 2.7-4.0 Kg 2.7-4.2 Kg 3-4 Kg 4.0-4.5 Kg 4.2-7.0 Kg 4.5-5.5 Kg
81 258 94 805 1637 523 5382
5-6 Kg 5.5-6.3 Kg 5.6-6.0 Kg 6-7 Kg 7-8 Kg
150 703 298 1322 273
```

```
mean(datos$Peso)
```

```
[1] 5.068799
```

**Se cuantifica un total de 10898 piezas premium, 557 categoria Standard y 71 piezas como industrial B.**

En tanto para los calibres se observa que 5328 piezas corresponden a calibre 4.5-5.5 Kg. Y por ultimo el peso promedio de este lote fue de 5.068799 Kg. ““