

# Clasificación de Salmón del Atlantico destinado a filete

Diplomado de Análisis de datos con R para la Acuicultura

Cristian Naguian Asenjo

20 May 2022

## Tipo de datos

Datos a analizar a partir de un TXT, las que corresponden a las piezas con destino a filete, a continuación se observa que tipo de variables se analizaran y su característica.

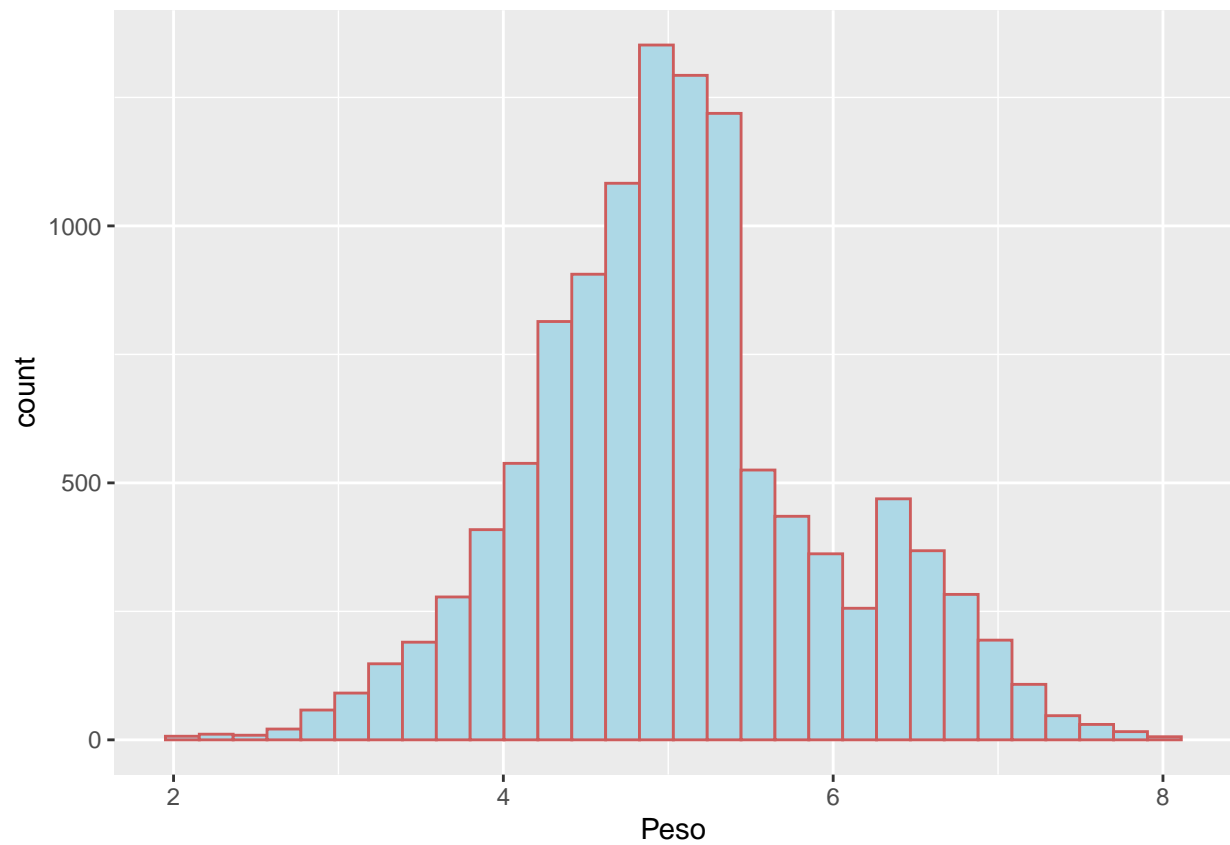
```
set.seed(1)
datos <- read.delim("/cloud/project/Piezas a filete.txt", na="NA")
str(datos)
```

```
## '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" "2.7-4.0" "2.7-4.0" "2.7-4.0" ...
## $ Calidad: chr   "Premium" "Premium" "Premium" "Premium" ...
```

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

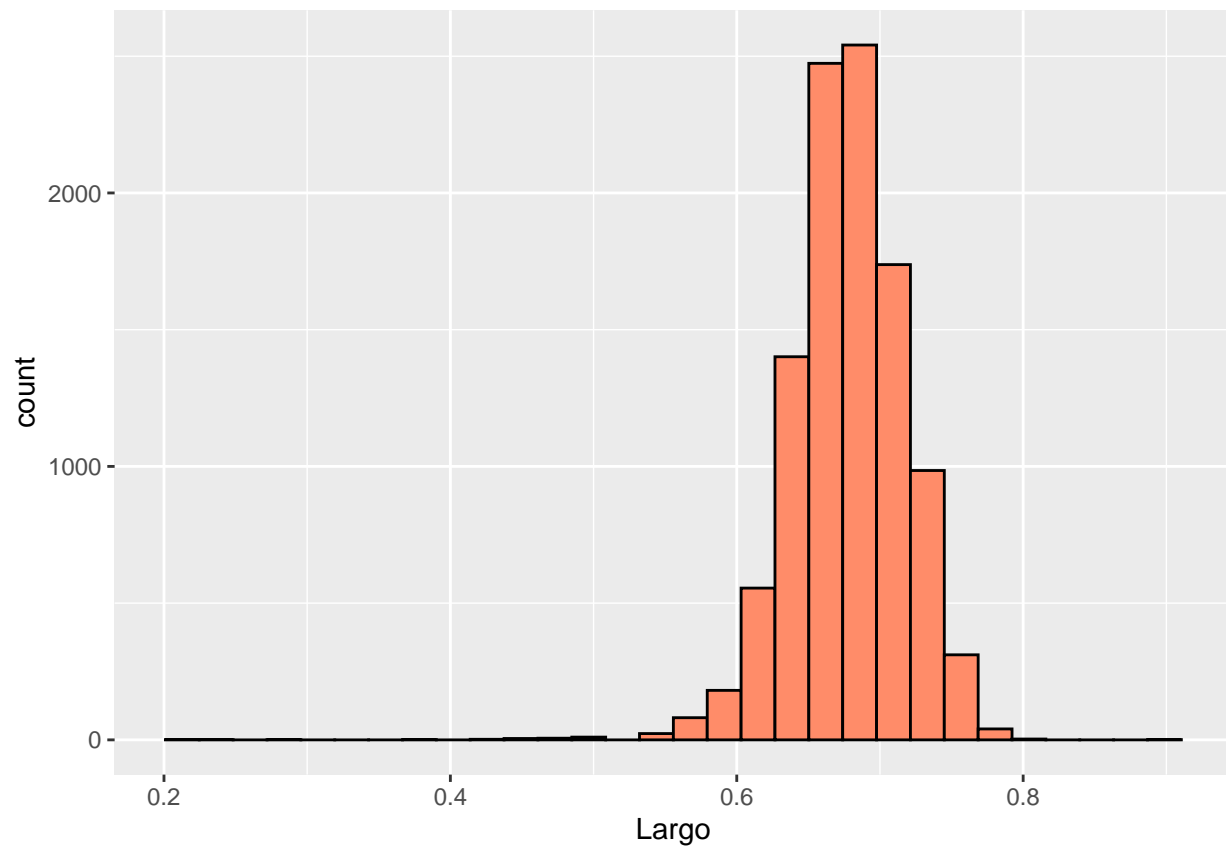
A continuación se observa dos histogramas para peso y largo, a partir de grafico con 30 barras de distribución, despues tenemos un gráfico de densidad el cual visualiza la distribución de datos cuantitativos para el peso en un intervalo o período de tiempo continuo. Los graficos de distribucion empirica acumulada se puede concluir que presentan una distribucion de tipo normal. (Largo y peso)

```
ggplot(datos, aes(x = Peso)) +geom_histogram(bins = 30, color = "indianred", fill="lightblue")
```



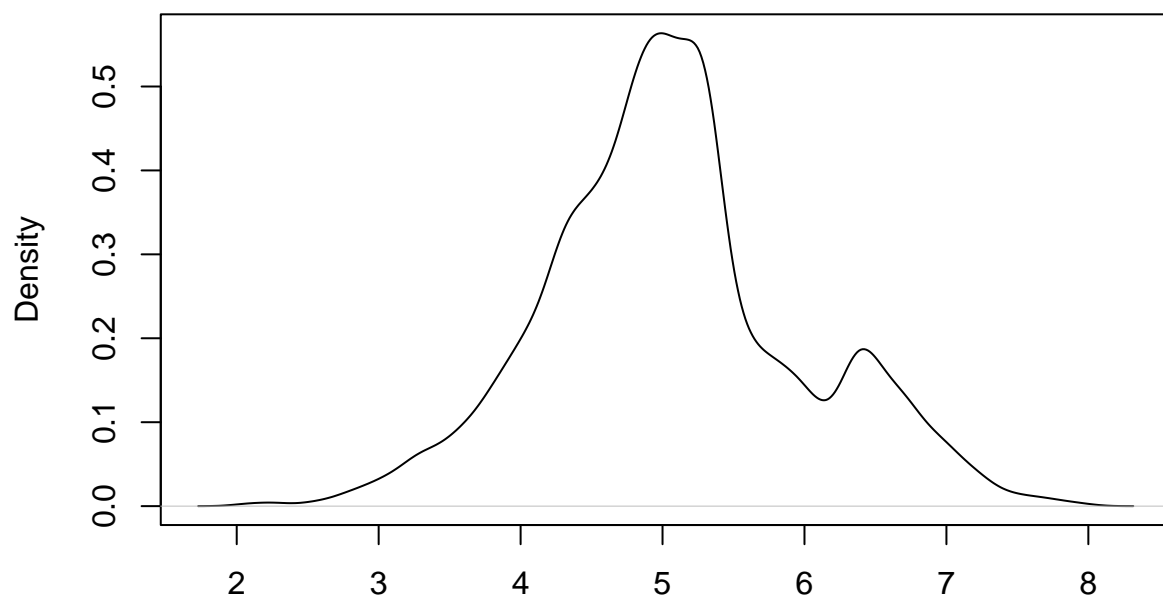
```
ggplot(datos, aes(x = Largo)) +geom_histogram(bins = 30, color = "black", fill="salmon1")
```

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



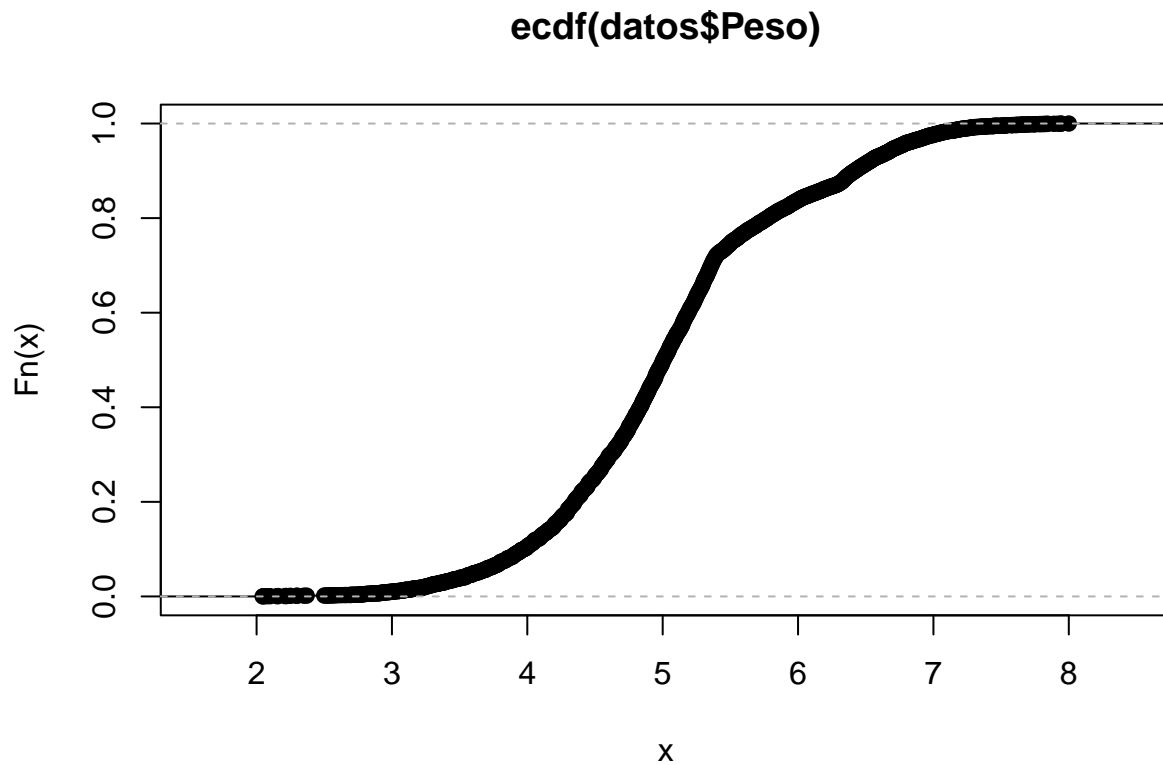
```
plot(density(datos$Peso))
```

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



N = 11526 Bandwidth = 0.1055

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



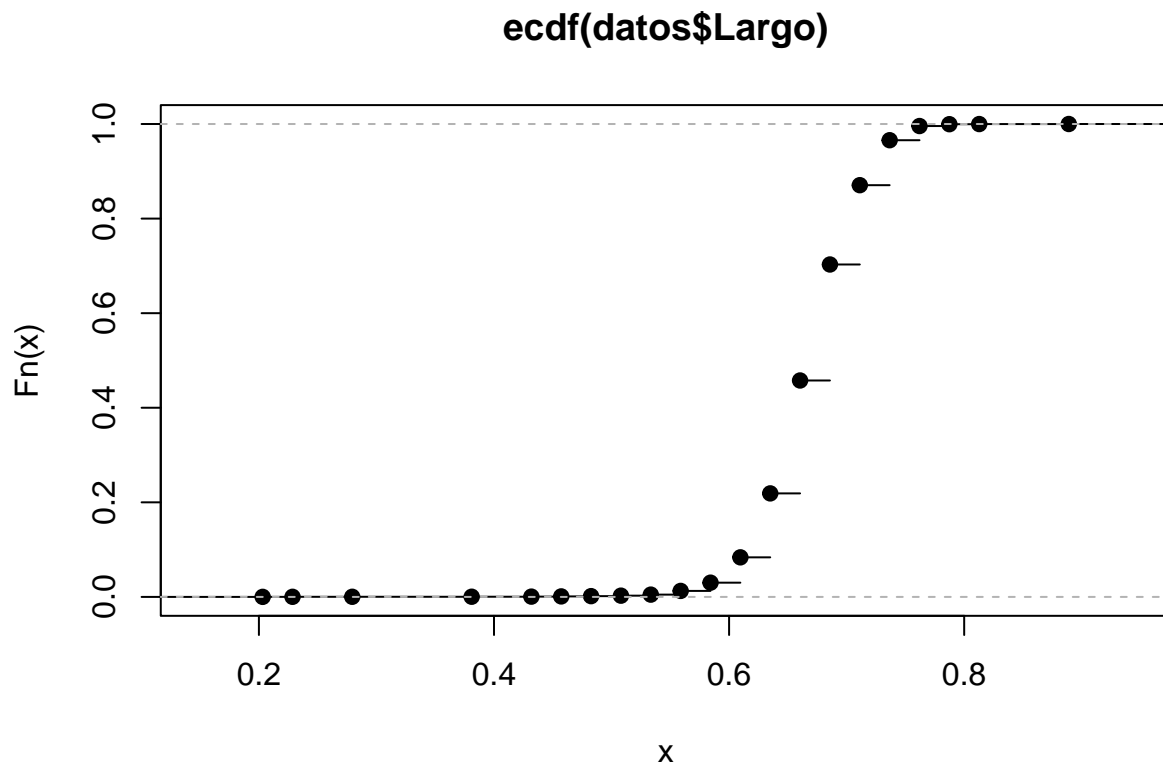
```
ecdf(datos$Peso) #Distribución empírica acumulada de la variable weight.
```

```
## Empirical CDF
```

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

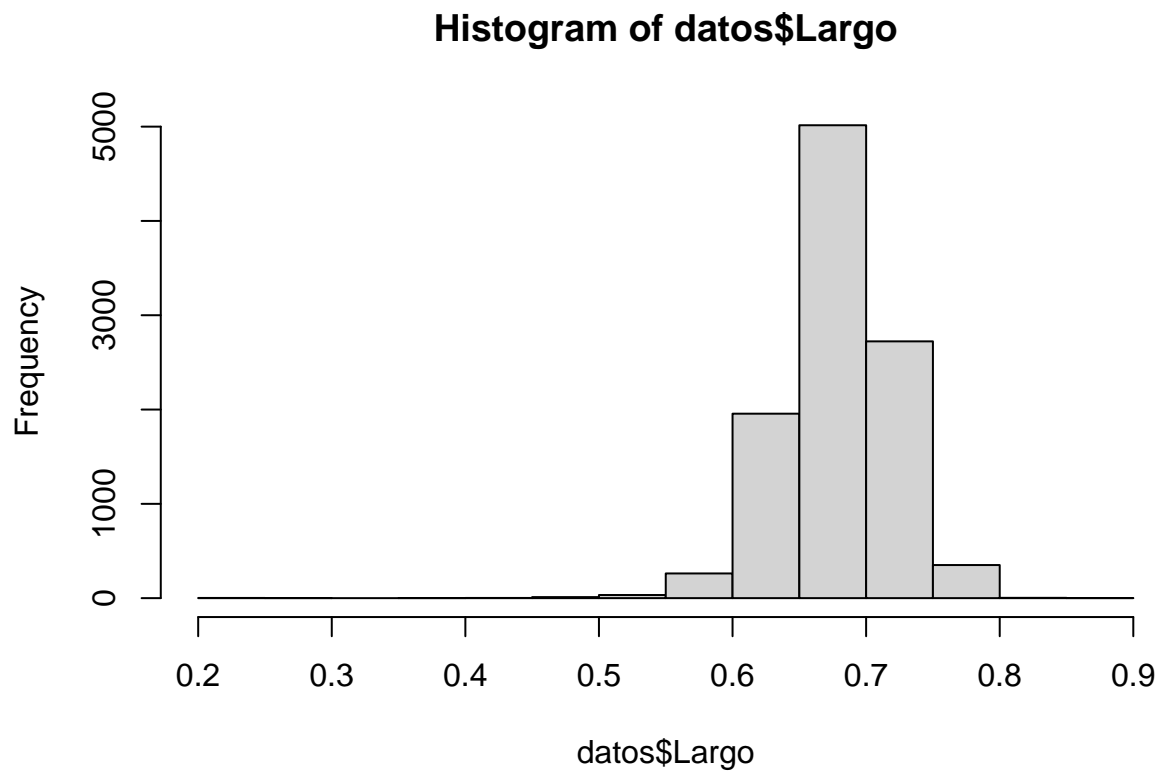
```
## x[1:982] = 2.045, 2.05, 2.075, ..., 7.945, 8
```

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

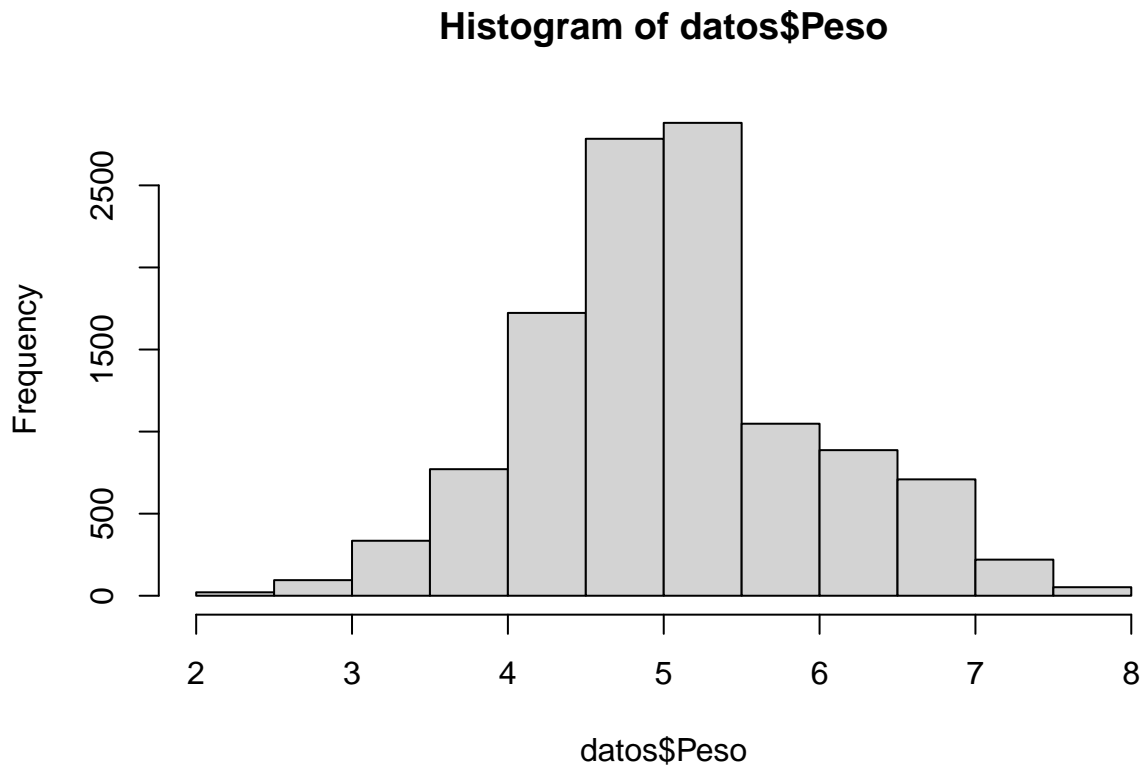


Identifica si los datos están balanceados o no entre tratamientos usando tablas de frecuencia

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



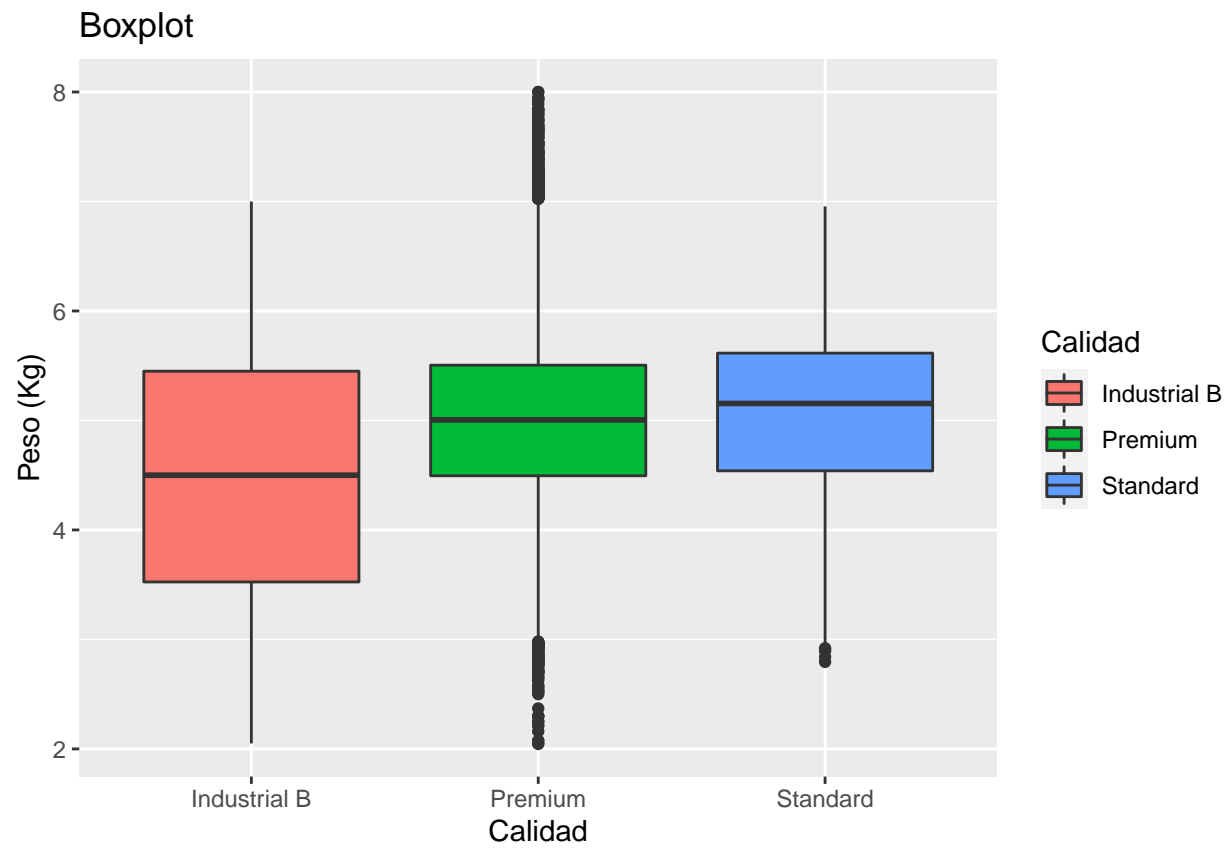
```
hist(datos$Peso)
```



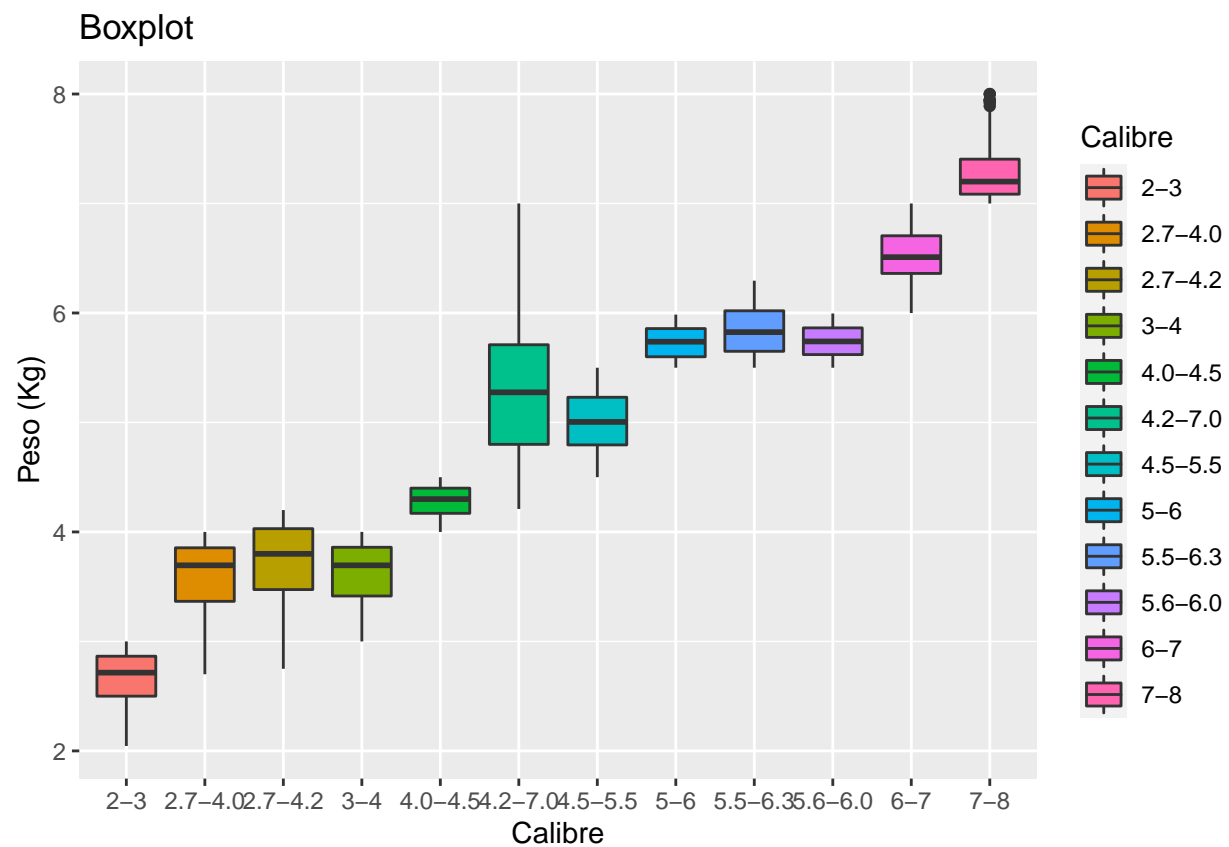
**Establece relación entre variables cuantitativas y factores usando gráficas de correlación, boxplot, interacción o de tamaño de los efectos**

Como se puede observar en los siguientes graficos de cajas , entre calidad peso y calibres, peso, los datos obtenidos para este lote no presentan una desviacion significativa. Se puede apreciar que los largos de las piezas para calidad Industrial B, no fueron medidas.

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



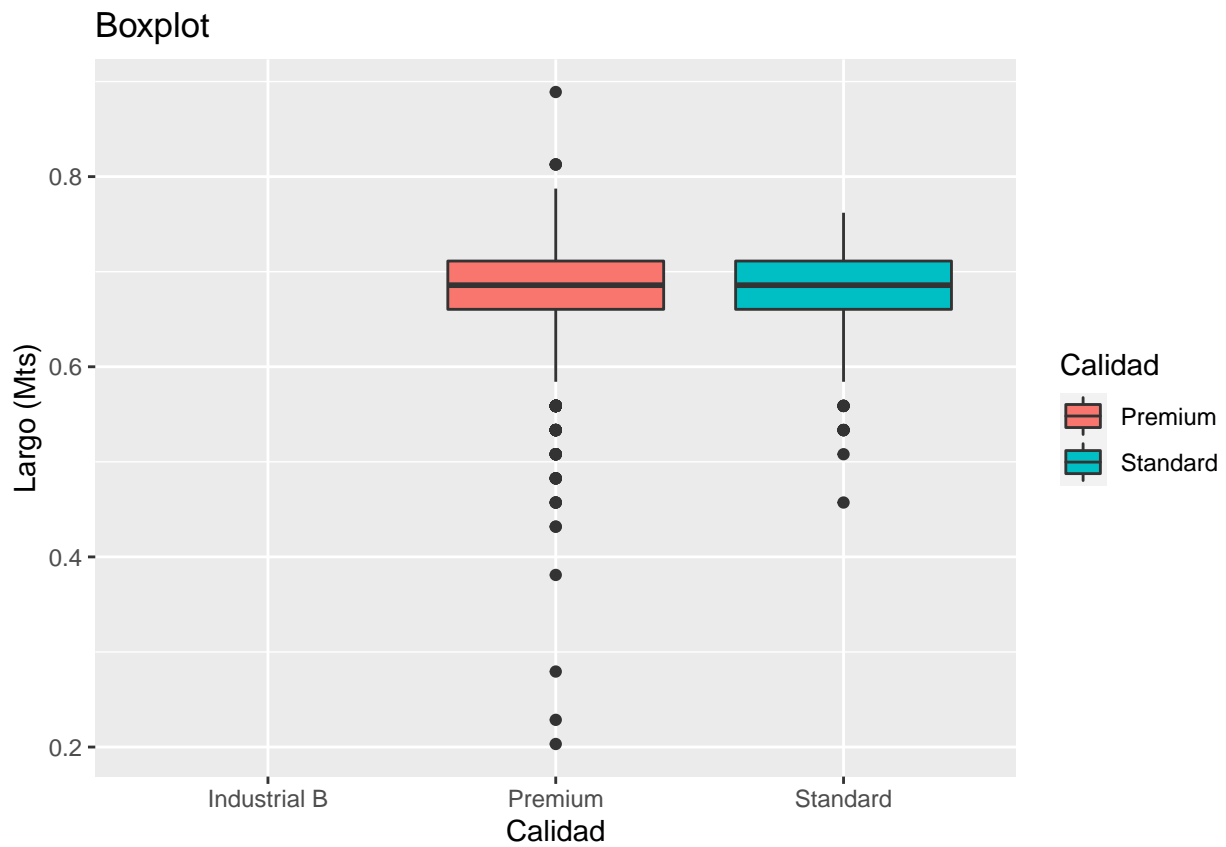
```
ggplot(datos, aes(x=Calibre, y=Peso, fill = Calibre)) +geom_boxplot()+labs(title="Boxplot", x="Calibre")
```



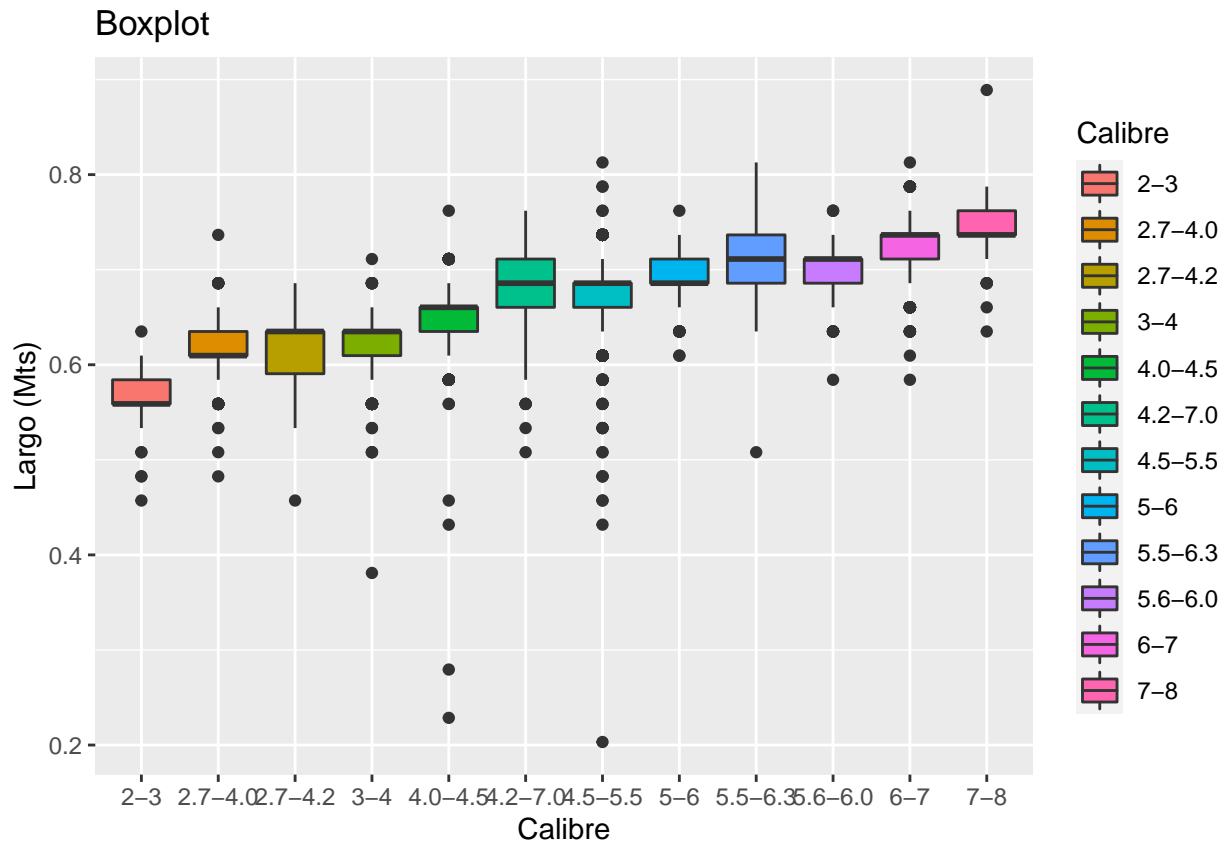
```
ggplot(datos, aes(x=Calidad, y=Largo, fill = Calidad)) +geom_boxplot()+labs(title="Boxplot", x="Calidad")
```

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





```
ggplot(datos, aes(x=Calibre, y=Largo, fill = Calibre)) +geom_boxplot()+labs(title="Boxplot", x="Calibre")  
## Warning: Removed 1165 rows containing non-finite values (stat_boxplot).
```



Identifica si existen errores, datos faltantes o valores atípicos

```
datos$Calibre <- as.factor(datos$Calibre)
datos$Calidad <- as.factor(datos$Calidad)
summary(datos)
```

```
##      Pieza      Peso      Largo      Calibre
## Min.   :    1  Min.   :2.045  Min.   :0.2032  4.5-5.5:5382
## 1st Qu.: 2882  1st Qu.:4.495  1st Qu.:0.6604  4.0-4.5:1637
## Median : 5764  Median :5.010  Median :0.6858  6-7    :1322
## Mean   : 5764  Mean   :5.069  Mean   :0.6769  3-4    : 805
## 3rd Qu.: 8645  3rd Qu.:5.515  3rd Qu.:0.7112  5.5-6.3: 703
## Max.   :11526  Max.   :8.000  Max.   :0.8890  4.2-7.0: 523
##                                     NA's   :1165  (Other):1154
##
##      Calidad
## Industrial B:   71
## Premium     :10898
## Standard    :  557
##
##
##
##
```

Como se puede observar en el resumen de datos, existen 1156 datos de largo los cuales no estan ingresados, debido a que la grader, máquina que calibra las piezas no pudo 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 2.7-4.0 2.7-4.2      3-4 4.0-4.5 4.2-7.0 4.5-5.5      5-6 5.5-6.3 5.6-6.0  
##      81      258      94      805      1637      523      5382      150      703      298  
##      6-7      7-8  
##     1322      273
```

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

```
## [1] 5.068799
```

```
mean(datos$Largo)
```

```
## [1] NA
```

```
sd(datos$Peso)
```

```
## [1] 0.9123698
```

```
summarise(datos)
```

```
## data frame with 0 columns and 1 row
```

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 último el peso promedio de este lote fue de 5.068799 Kg.f ““