

# QTL-SRS

Marcos Mancilla

02 November 2021

## Realizando análisis exploratorio de datos

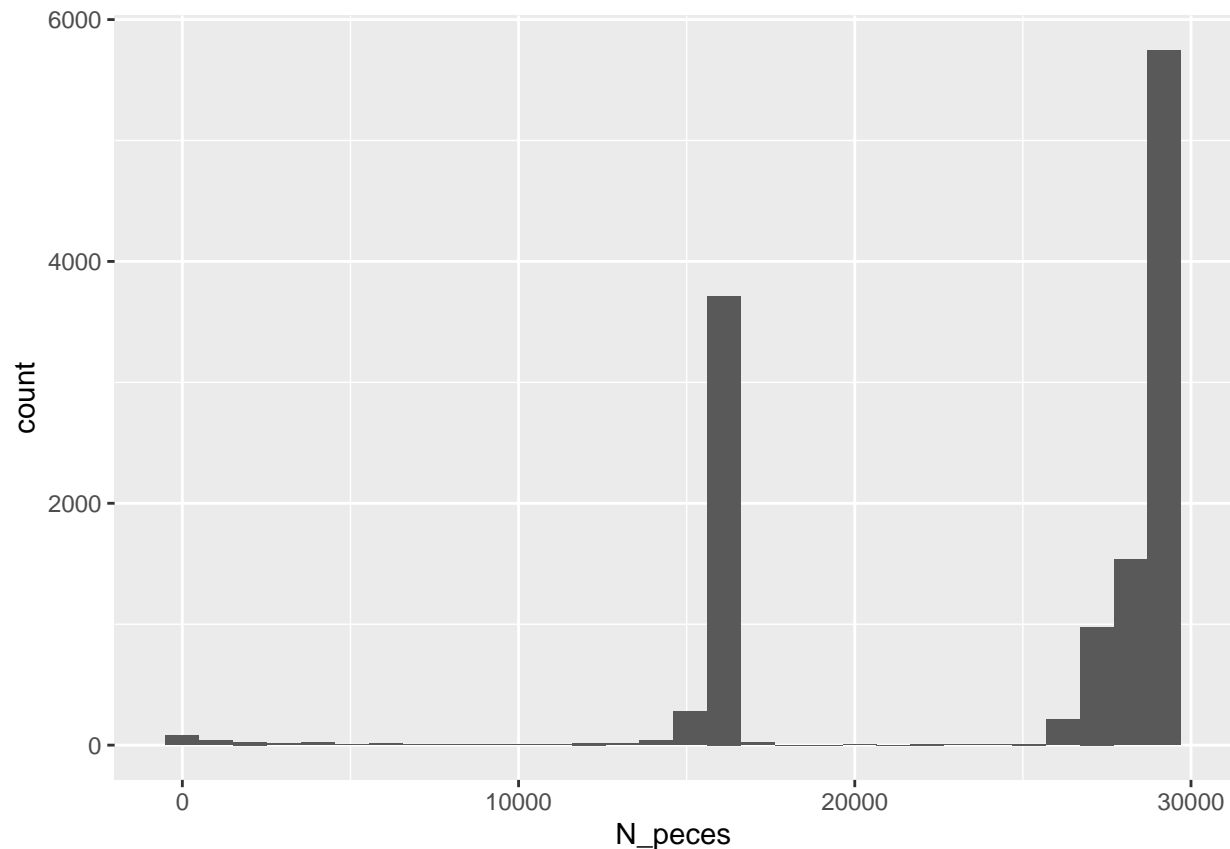
```
##      Fecha                Centro                Unidad
##  Min.   :2019-02-09 00:00:00  Length:12825      Length:12825
##  1st Qu.:2019-06-01 00:00:00  Class :character  Class :character
##  Median :2019-07-26 00:00:00  Mode  :character  Mode  :character
##  Mean   :2019-07-25 16:53:46
##  3rd Qu.:2019-09-20 00:00:00
##  Max.   :2019-12-09 00:00:00
##      Subgrupo      N_peces      Biomasa_kg      Mortalidad
##  Length:12825      Min.   :    0      Min.   :    0      Min.   :    0.000
##  Class :character  1st Qu.:16308      1st Qu.: 23766      1st Qu.:    0.000
##  Mode  :character  Median :28340      Median : 42046      Median :    1.000
##                      Mean   :24143      Mean   : 44360      Mean   :    9.889
##                      3rd Qu.:29028      3rd Qu.: 59946      3rd Qu.:    8.000
##                      Max.   :29209      Max.   :114599      Max.   :1982.000
##  Mortalidad_Biomasa_kg  N_cosecha      Biomasa_cosecha      Alimento_kg
##  Min.   :    0.000      Min.   :    0.0      Min.   :    0.0      Min.   :    0.0
##  1st Qu.:    0.000      1st Qu.:    0.0      1st Qu.:    0.0      1st Qu.: 285.0
##  Median :    2.015      Median :    0.0      Median :    0.0      Median : 446.0
##  Mean   :   27.549      Mean   :   76.9      Mean   : 270.1      Mean   : 440.6
##  3rd Qu.:   17.845      3rd Qu.:    0.0      3rd Qu.:    0.0      3rd Qu.: 604.0
##  Max.   :7671.965      Max.   :22602.0      Max.   :78009.3      Max.   :1560.0
##  Temperatura
##  Min.   : 0.00
##  1st Qu.:10.79
##  Median :11.20
##  Mean   :11.15
##  3rd Qu.:11.70
##  Max.   :15.00
## # A tibble: 6 x 12
##   Fecha                Centro Unidad Subgrupo N_peces Biomasa_kg Mortalidad
##   <dtm>                <chr> <chr> <chr>      <dbl>      <dbl>      <dbl>
## 1 2019-08-24 00:00:00 A      205   QTL2      16302      35625.      0
## 2 2019-08-25 00:00:00 A      205   QTL2      16302      35879.      0
## 3 2019-08-26 00:00:00 A      205   QTL2      16302      36261.      0
## 4 2019-08-27 00:00:00 A      205   QTL2      16302      36684.      0
## 5 2019-08-28 00:00:00 A      205   QTL2      16302      37100.      0
## 6 2019-08-29 00:00:00 A      205   QTL2      16302      37435.      0
## # ... with 5 more variables: Mortalidad_Biomasa_kg <dbl>, N_cosecha <dbl>,
## #   Biomasa_cosecha <dbl>, Alimento_kg <dbl>, Temperatura <dbl>
```

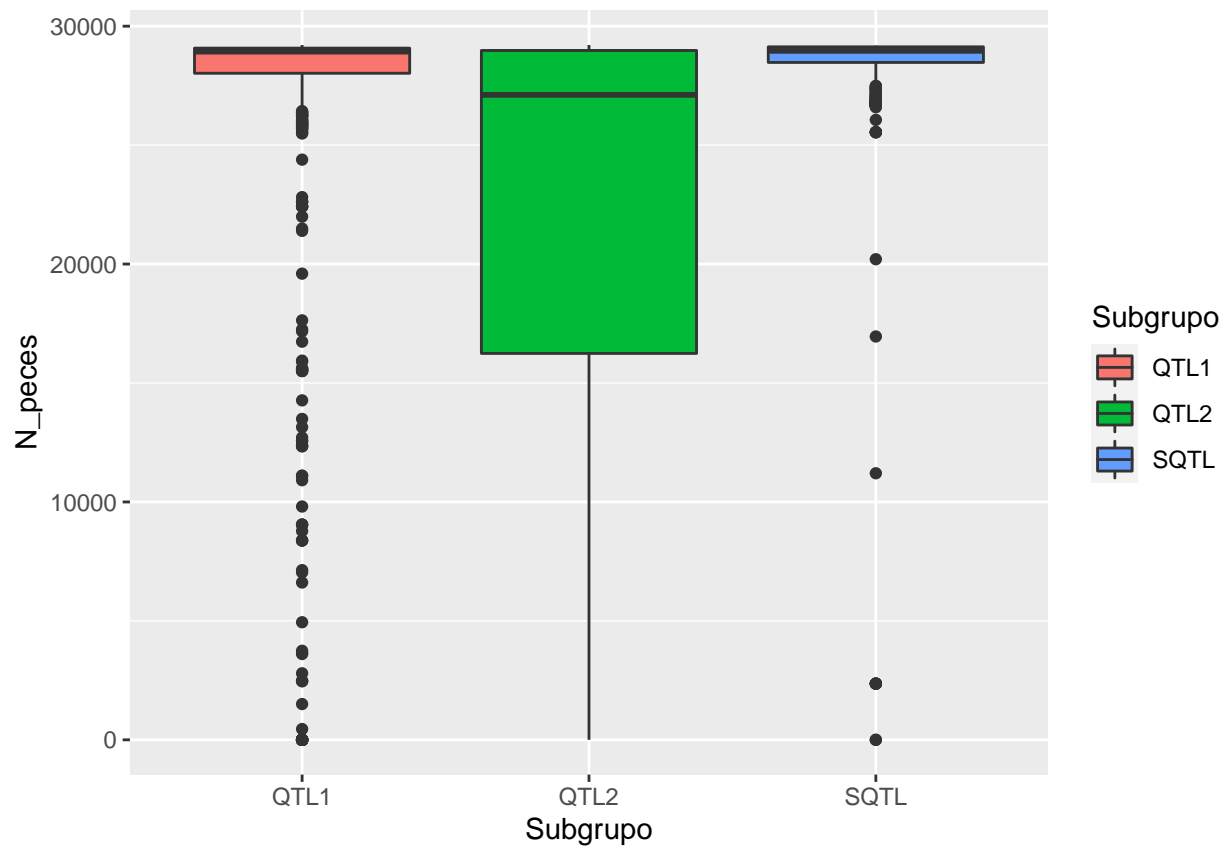
```
## tibble [12,825 x 12] (S3: tbl_df/tbl/data.frame)
## $ Fecha          : POSIXct[1:12825], format: "2019-08-24" "2019-08-25" ...
## $ Centro         : chr [1:12825] "A" "A" "A" "A" ...
## $ Unidad         : chr [1:12825] "205" "205" "205" "205" ...
## $ Subgrupo       : chr [1:12825] "QTL2" "QTL2" "QTL2" "QTL2" ...
## $ N_peces        : num [1:12825] 16302 16302 16302 16302 16302 ...
## $ Biomasa_kg     : num [1:12825] 35625 35879 36261 36684 37100 ...
## $ Mortalidad     : num [1:12825] 0 0 0 0 0 0 0 1 2 2 ...
## $ Mortalidad_Biomasa_kg: num [1:12825] 0 0 0 0 0 ...
## $ N_cosecha      : num [1:12825] 0 0 0 0 0 0 0 0 0 0 ...
## $ Biomasa_cosecha : num [1:12825] 0 0 0 0 0 0 0 0 0 0 ...
## $ Alimento_kg    : num [1:12825] 455 300 450 500 490 396 472 562 435 556 ...
## $ Temperatura   : num [1:12825] 11.3 11.1 11.1 11.3 10.3 ...
```

Observamos que el set de datos tiene 12.825 observaciones de 12 variables. Existe una variable fecha, 3 variables como texto (Centro, Unidad y Subgrupo) y 8 variables numéricas. Nos se aprecian datos faltantes.

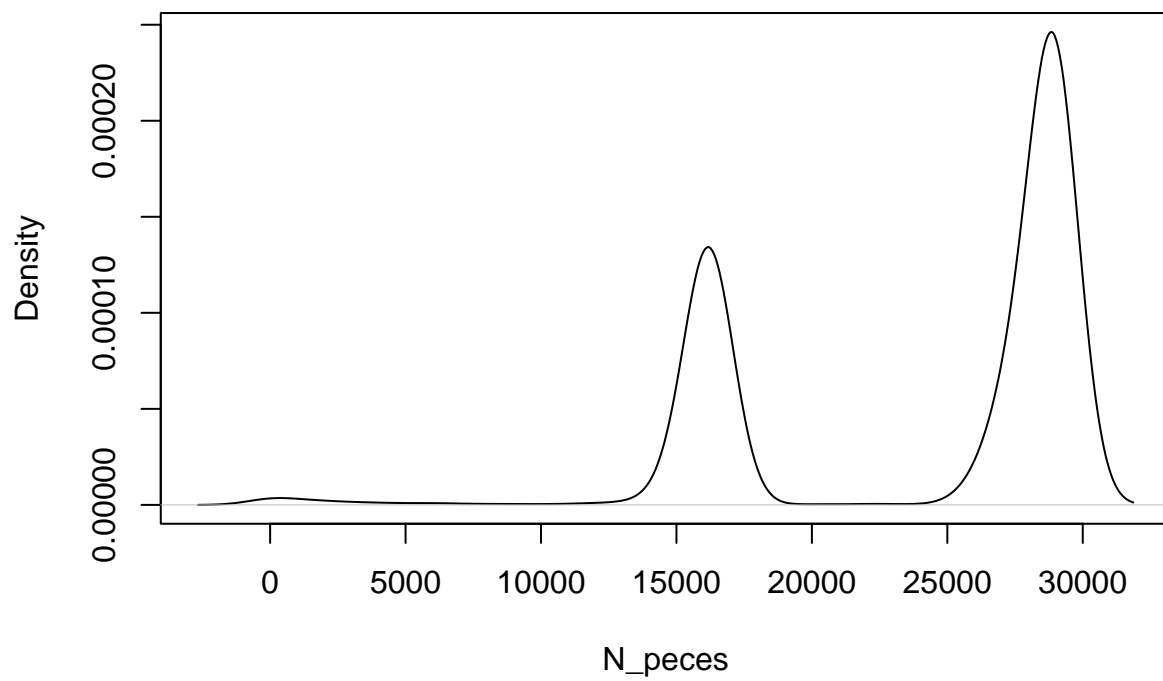
Graficamos variables de interés (histogramas, boxplots, densidad y densidad acumulada).

### Peces por subgrupo

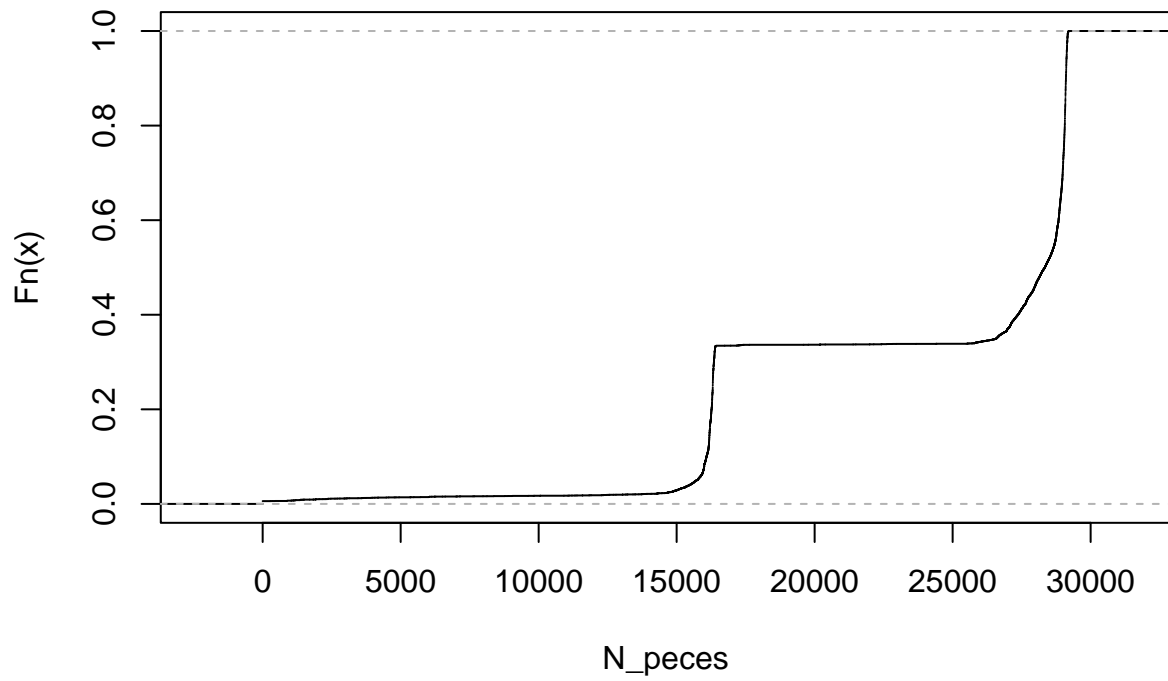




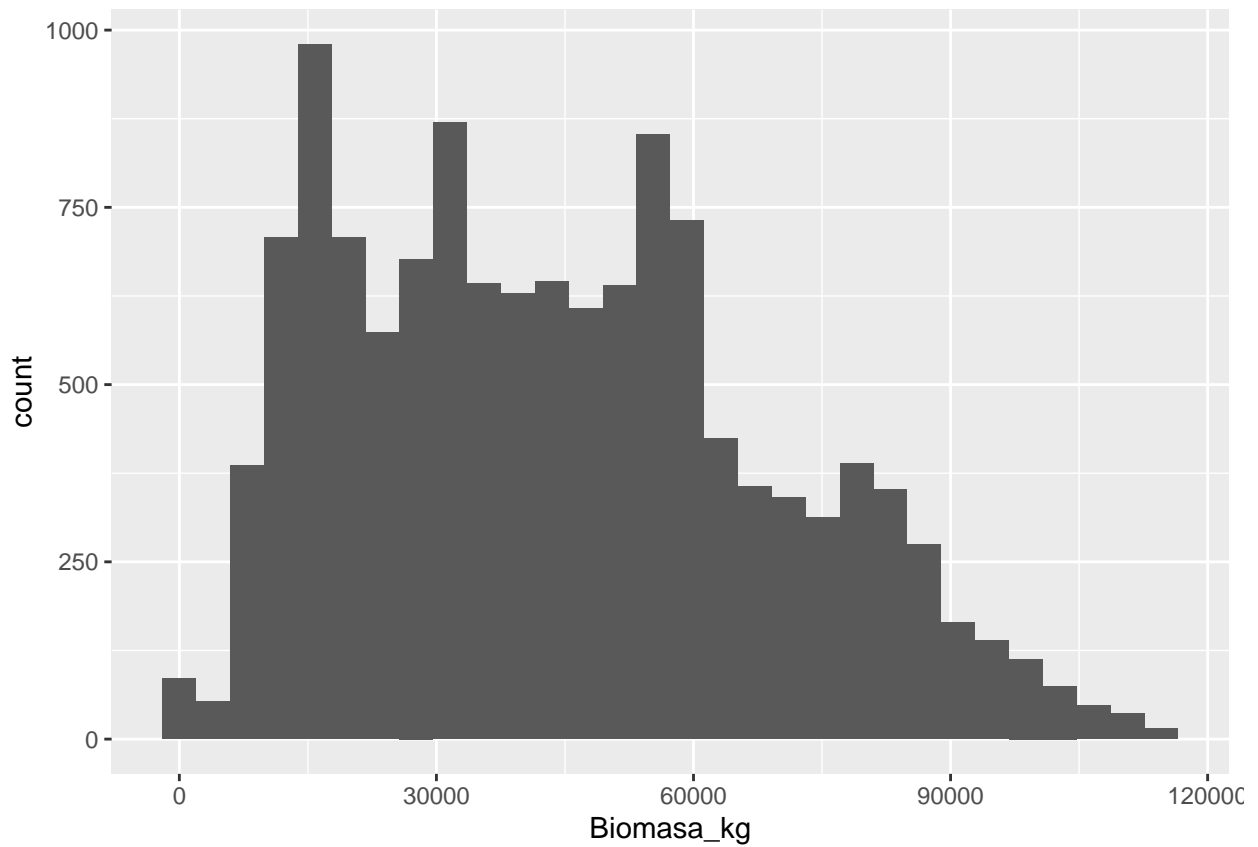
**Densidad empírica**

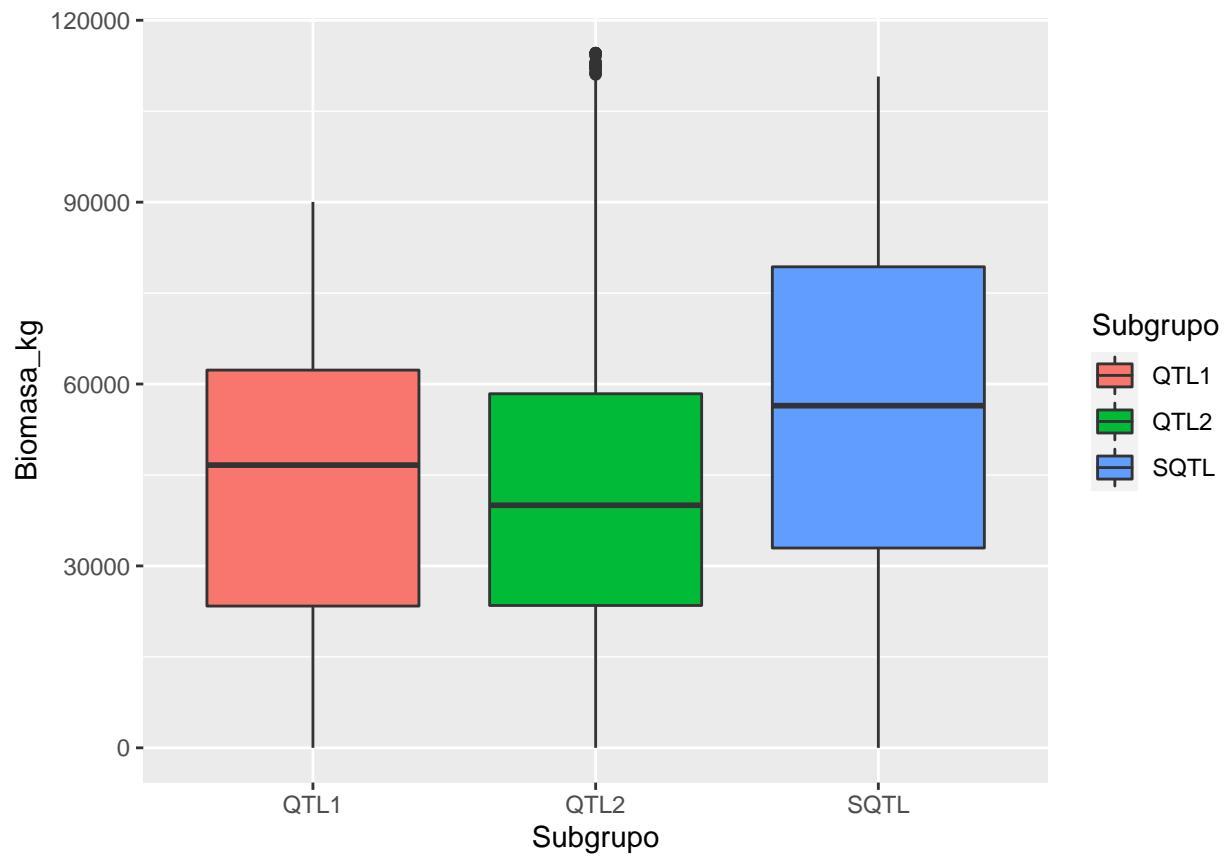


### Distribución acumulada empírica

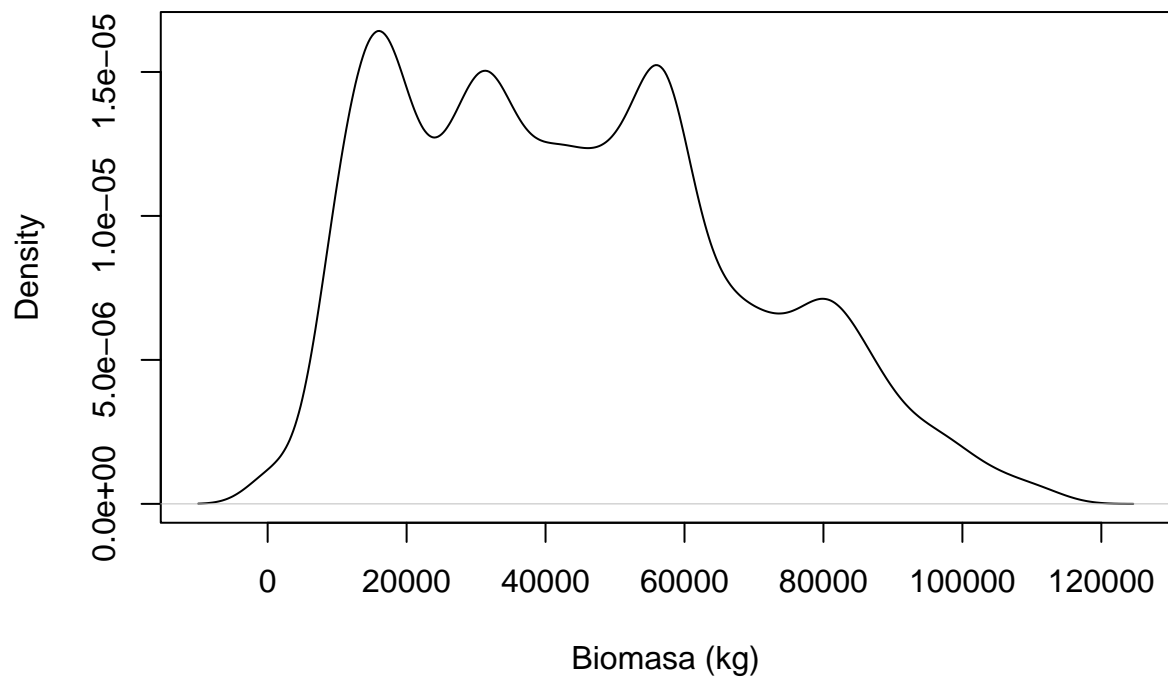


### Biomasa viva por subgrupo

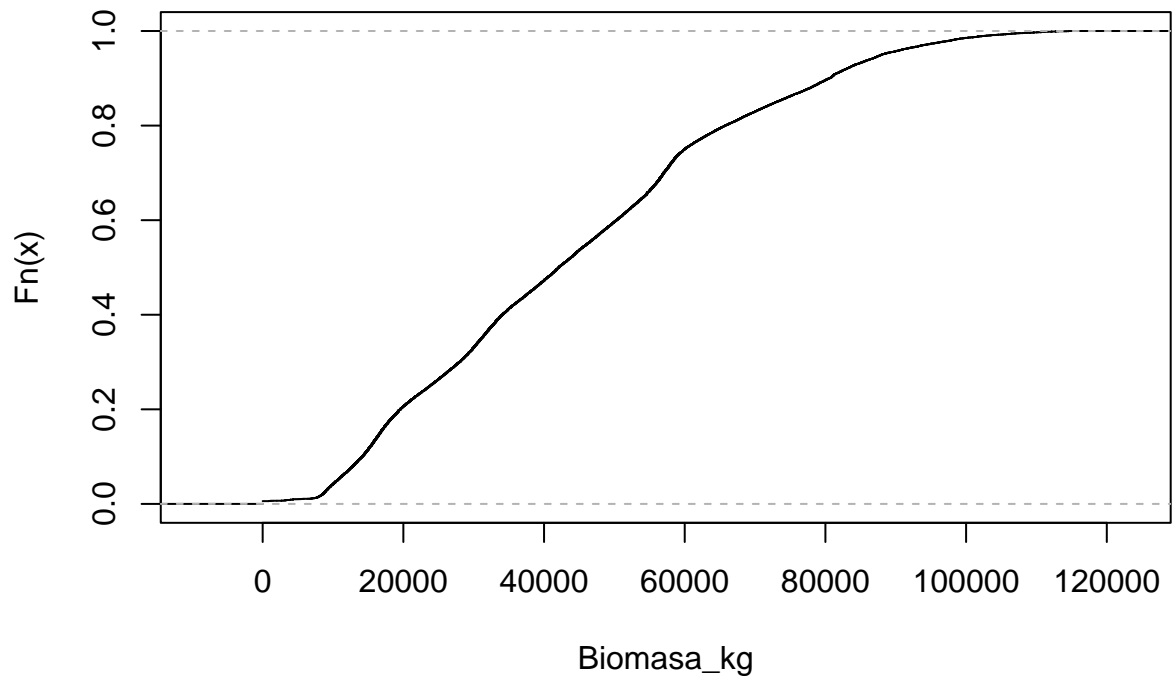




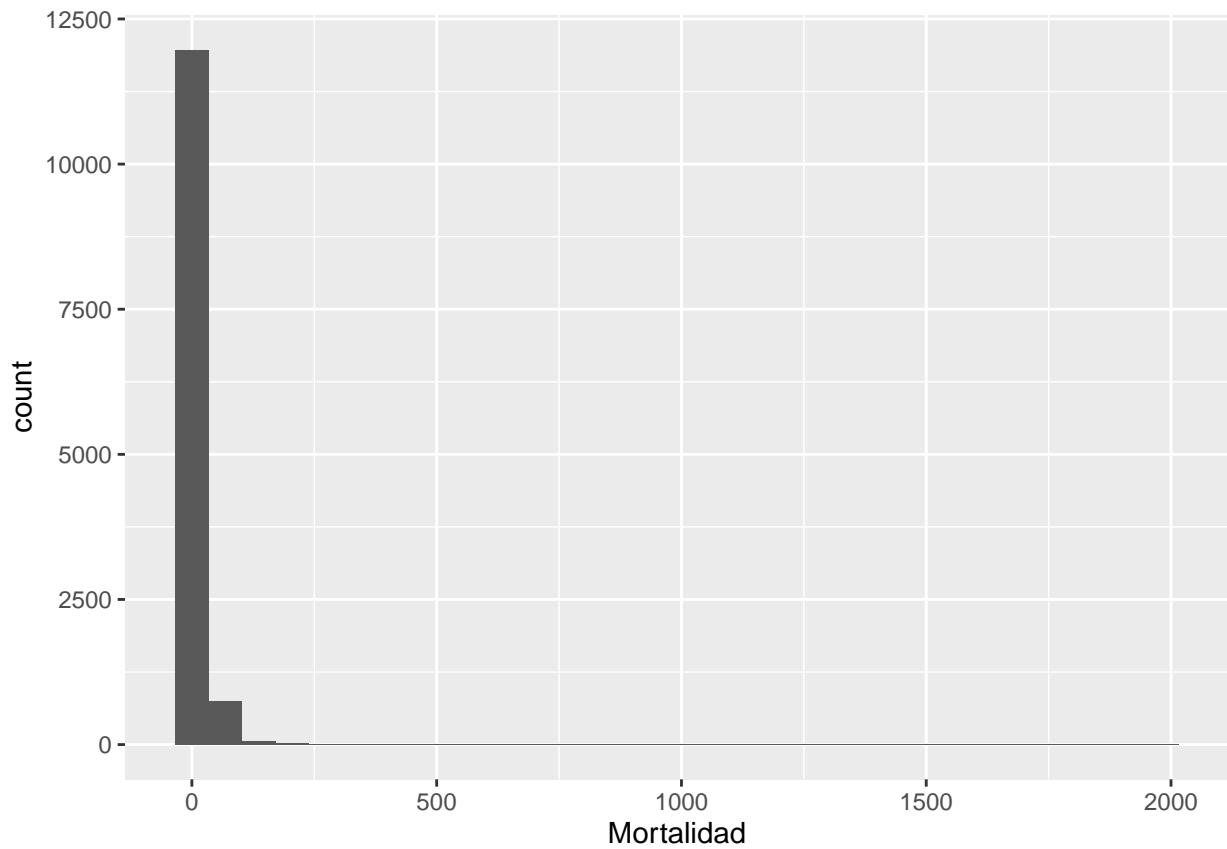
**Densidad empírica**



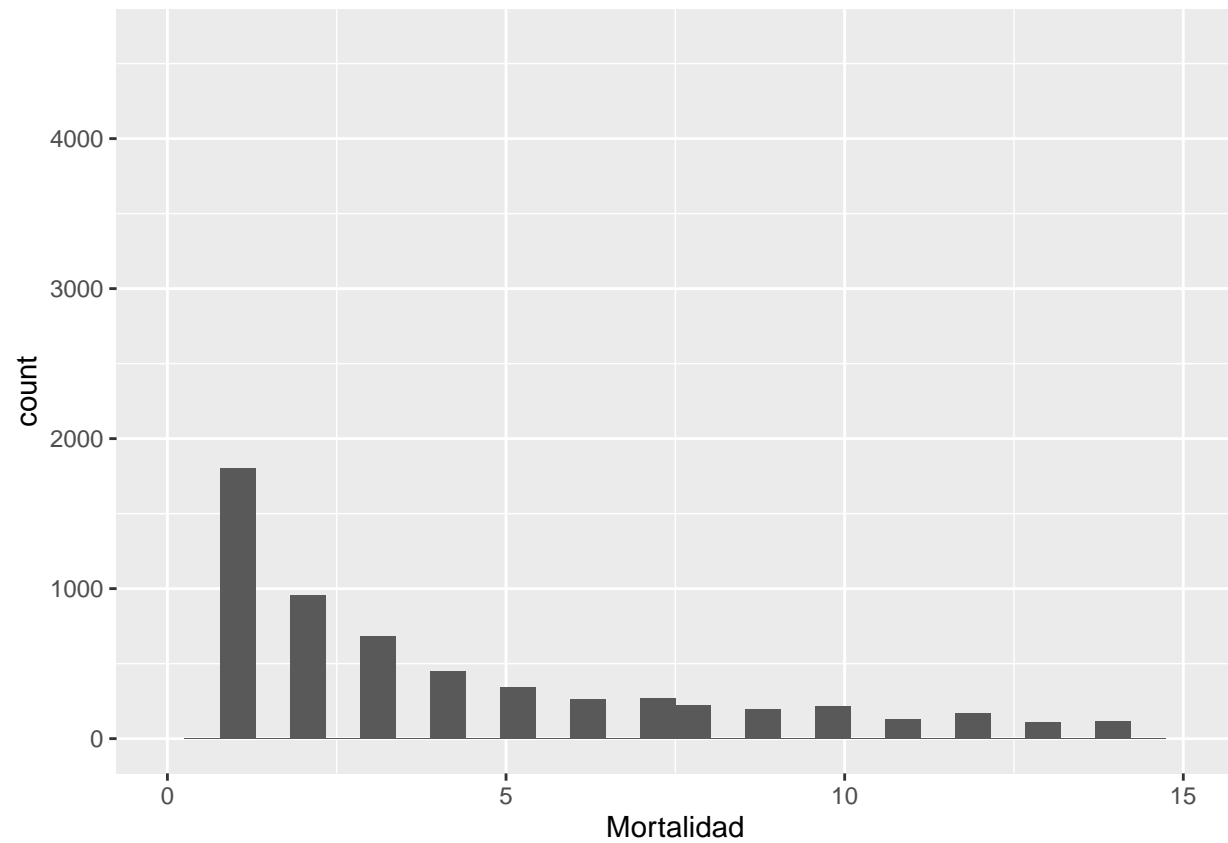
### Distribución acumulada empírica

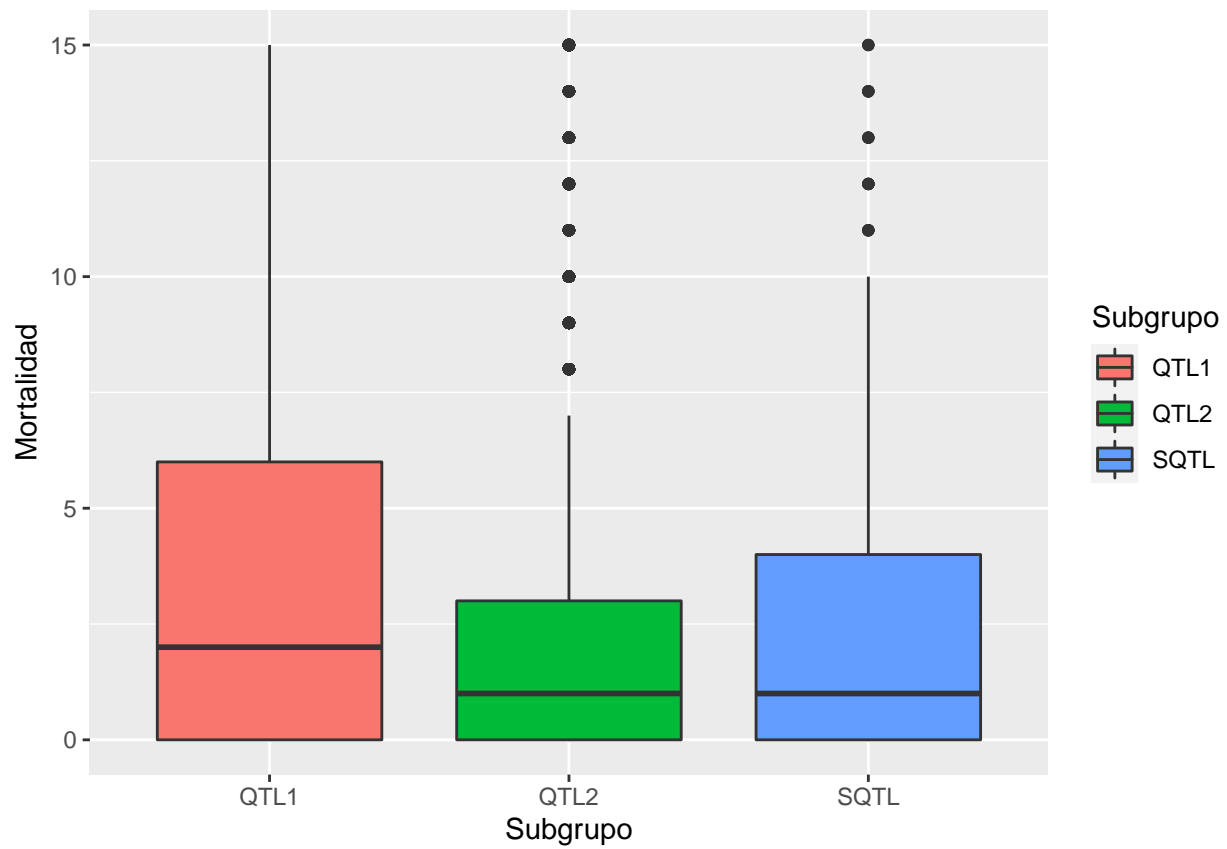


### Mortalidad (n° peces) por subgrupo

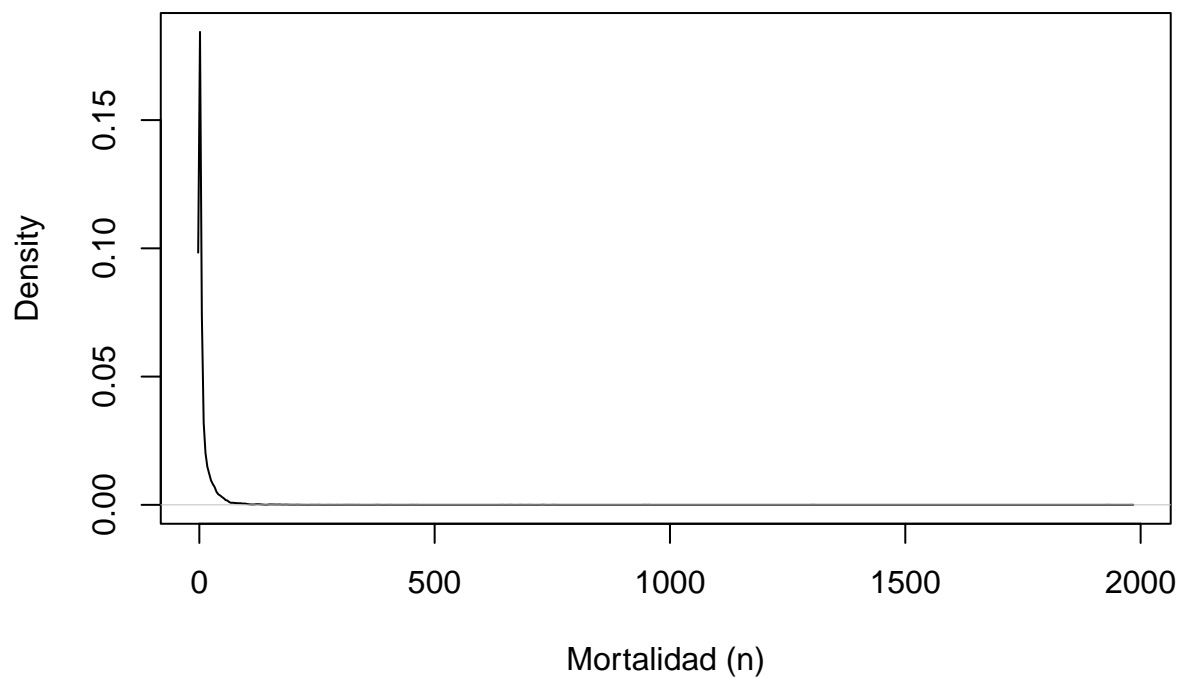


Para apreciar mejor como se comporta esta variable, haremos un histograma ajustando el eje x



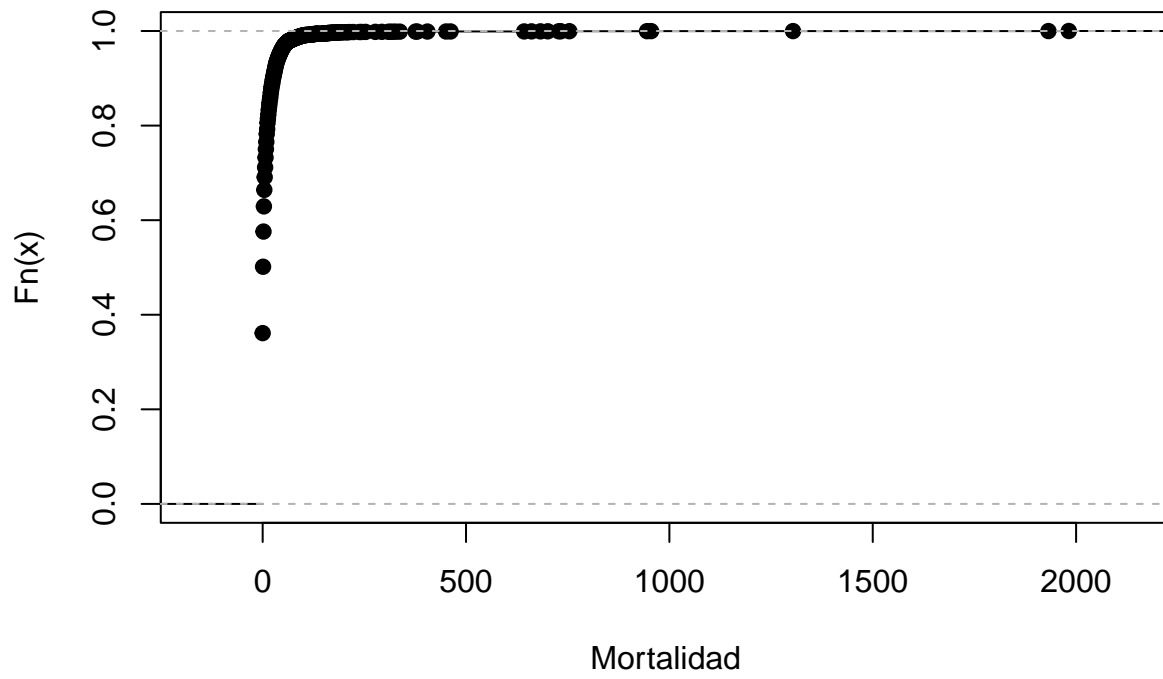


### Densidad empírica

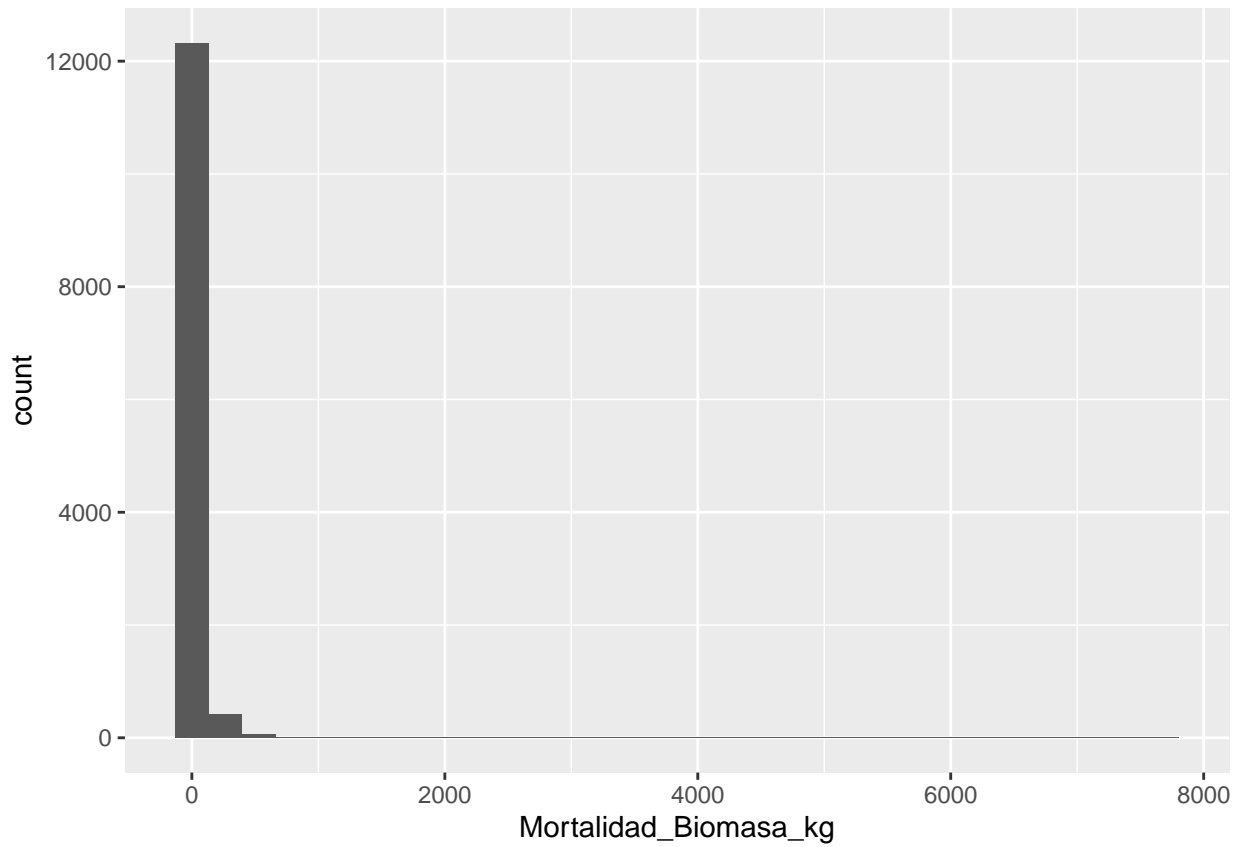


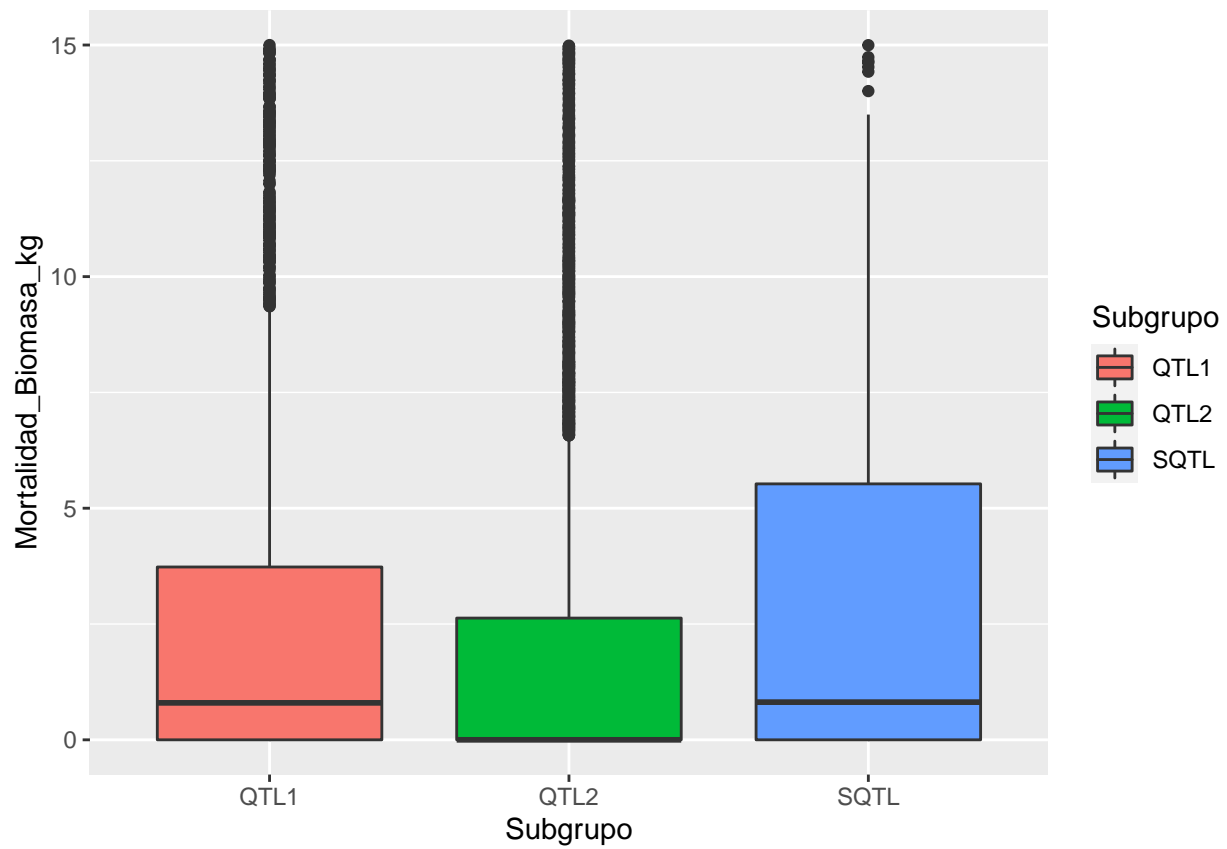


**Distribución acumulada empírica**

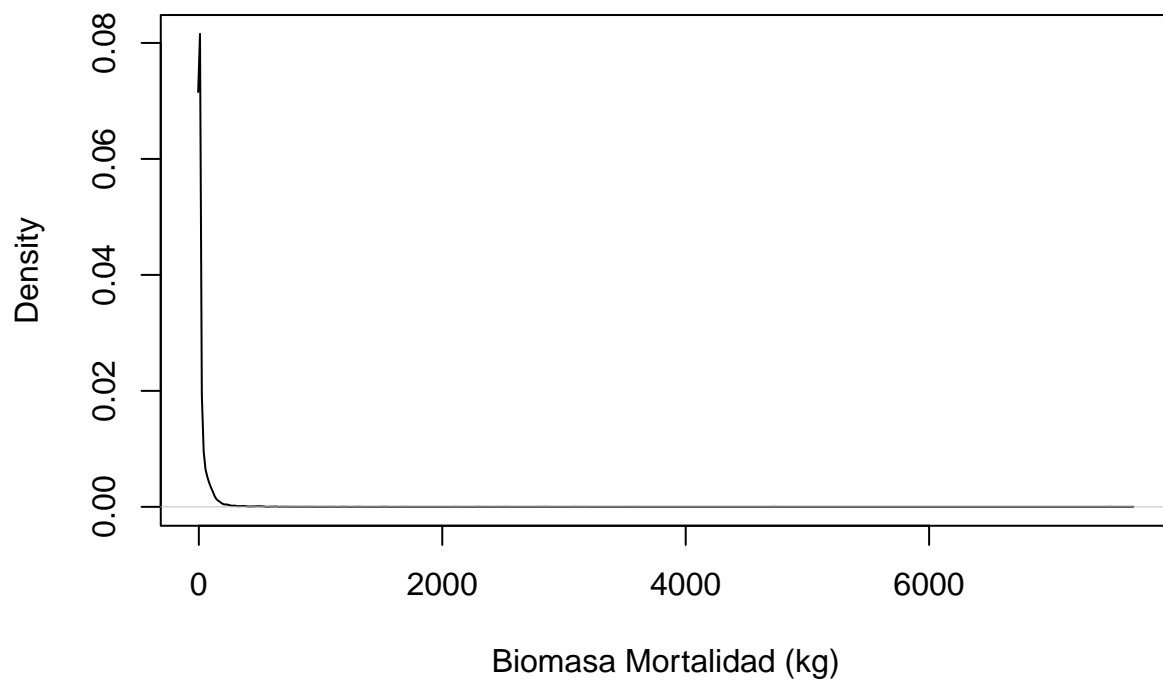


**Biomasa de la mortalidad por subgrupo**

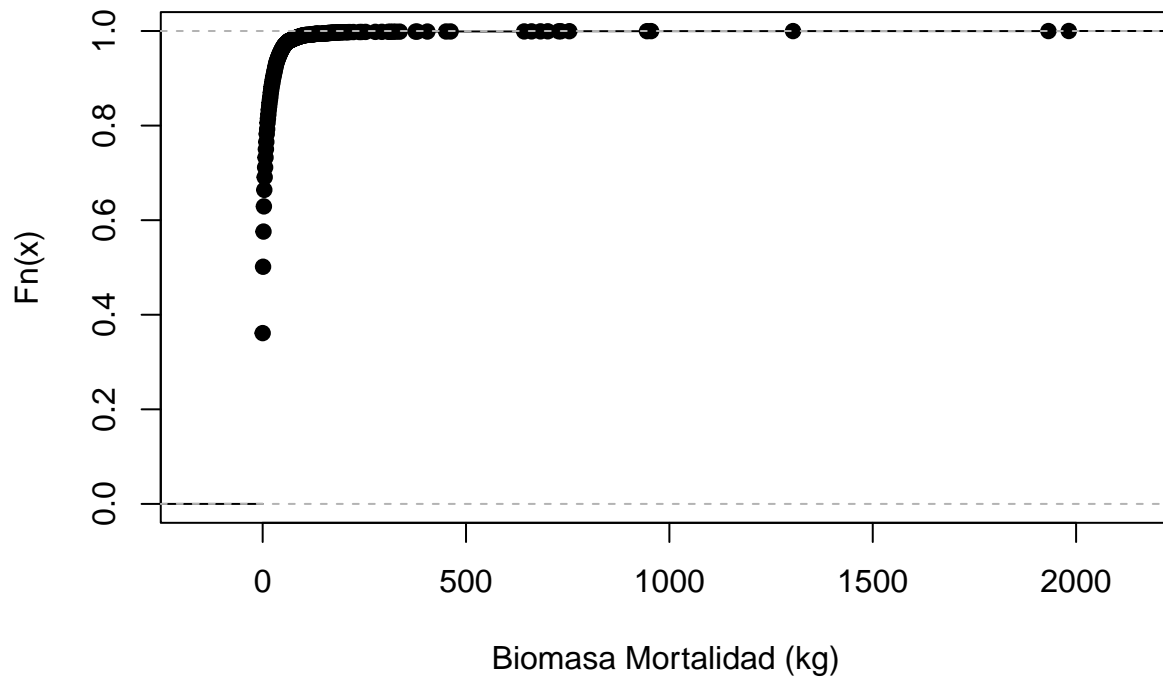




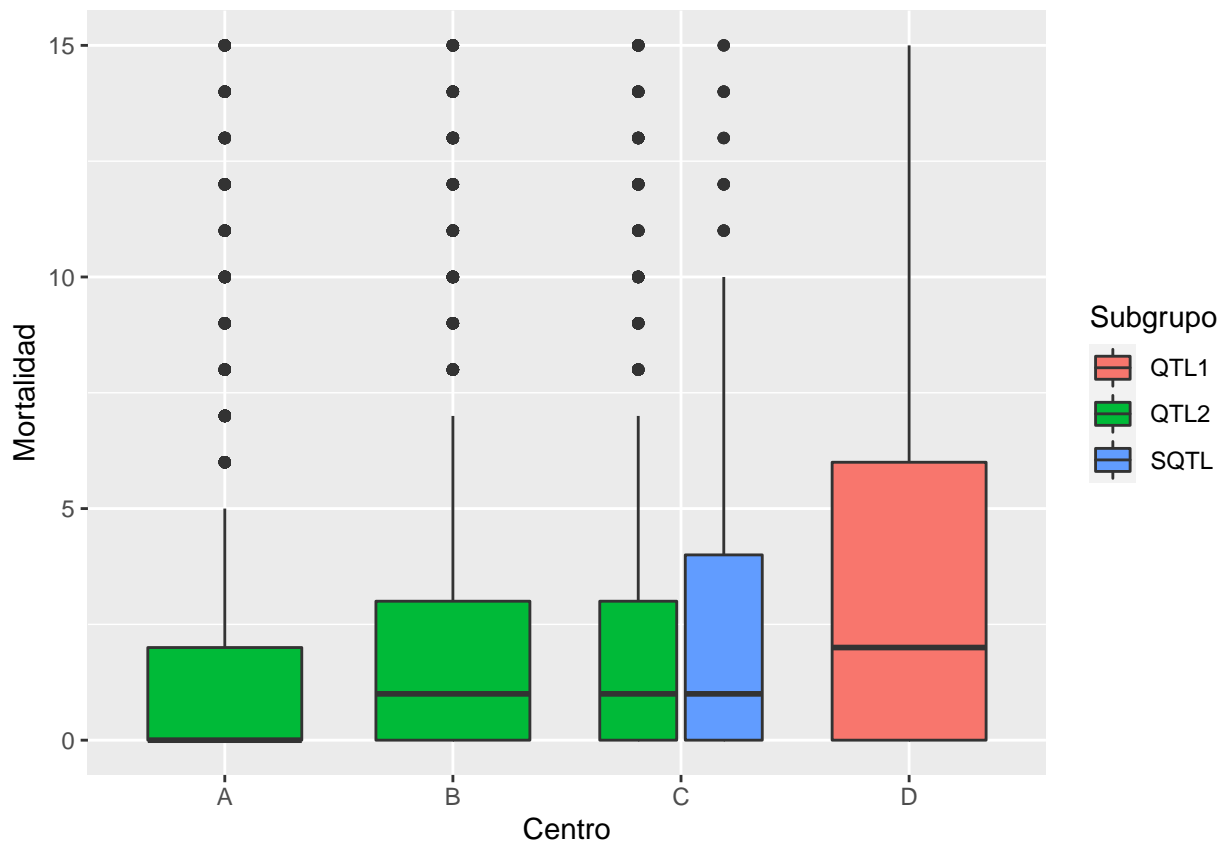
**Densidad empírica**



Distribución acumulada empírica



Hacemos boxplot de variable respuesta en función de otras variables de interés.



## Transformaciones de variables a factor

```
##
## QTL1 QTL2 SQTL
## 3256 9078 491

##
##   A    B    C    D
## 4130 3222 2217 3256

## tibble [12,825 x 12] (S3: tbl_df/tbl/data.frame)
##  $ Fecha           : POSIXct[1:12825], format: "2019-08-24" "2019-08-25" ...
##  $ Centro           : Factor w/ 4 levels "A","B","C","D": 1 1 1 1 1 1 1 1 1 1 ...
##  $ Unidad           : Factor w/ 24 levels "101","102","103",...: 19 19 19 19 19 19 19 19 19 19 ...
##  $ Subgrupo         : Factor w/ 3 levels "QTL1","QTL2",...: 2 2 2 2 2 2 2 2 2 2 ...
##  $ N_peces          : num [1:12825] 16302 16302 16302 16302 16302 ...
##  $ Biomasa_kg       : num [1:12825] 35625 35879 36261 36684 37100 ...
##  $ Mortalidad        : num [1:12825] 0 0 0 0 0 0 0 1 2 2 ...
##  $ Mortalidad_Biomasa_kg: num [1:12825] 0 0 0 0 0 ...
##  $ N_cosecha         : num [1:12825] 0 0 0 0 0 0 0 0 0 0 ...
##  $ Biomasa_cosecha   : num [1:12825] 0 0 0 0 0 0 0 0 0 0 ...
##  $ Alimento_kg       : num [1:12825] 455 300 450 500 490 396 472 562 435 556 ...
##  $ Temperatura       : num [1:12825] 11.3 11.1 11.1 11.3 10.3 ...
```

```

##      Fecha                Centro      Unidad      Subgrupo
## Min.   :2019-02-09 00:00:00 A:4130    105      : 909    QTL1:3256
## 1st Qu.:2019-06-01 00:00:00 B:3222    107      : 902    QTL2:9078
## Median :2019-07-26 00:00:00 C:2217    104      : 900    SQTL: 491
## Mean   :2019-07-25 16:53:46 D:3256    110      : 893
## 3rd Qu.:2019-09-20 00:00:00      106      : 890
## Max.   :2019-12-09 00:00:00      109      : 889
##                                     (Other):7442
##      N_peces      Biomasa_kg      Mortalidad      Mortalidad_Biomasa_kg
## Min.   :      0      Min.   :      0      Min.   :      0.000      Min.   :      0.000
## 1st Qu.:16308      1st Qu.: 23766      1st Qu.:      0.000      1st Qu.:      0.000
## Median :28340      Median : 42046      Median :      1.000      Median :      2.015
## Mean   :24143      Mean   : 44360      Mean   :      9.889      Mean   :     27.549
## 3rd Qu.:29028      3rd Qu.: 59946      3rd Qu.:      8.000      3rd Qu.:     17.845
## Max.   :29209      Max.   :114599      Max.   :1982.000      Max.   :7671.965
##
##      N_cosecha      Biomasa_cosecha      Alimento_kg      Temperatura
## Min.   :      0.0      Min.   :      0.0      Min.   :      0.0      Min.   :      0.00
## 1st Qu.:      0.0      1st Qu.:      0.0      1st Qu.: 285.0      1st Qu.:10.79
## Median :      0.0      Median :      0.0      Median : 446.0      Median :11.20
## Mean   :     76.9      Mean   :    270.1      Mean   : 440.6      Mean   :11.15
## 3rd Qu.:      0.0      3rd Qu.:      0.0      3rd Qu.: 604.0      3rd Qu.:11.70
## Max.   :22602.0      Max.   :78009.3      Max.   :1560.0      Max.   :15.00
##

```

Determinar si los datos están balanceados

Calculamos el número de observaciones por Subgrupo. Luego expresamos la cifra como proporción.

```

##
##      A      B      C      D
## QTL1      0      0      0 3256
## QTL2 4130 3222 1726      0
## SQTL      0      0 491      0
##
##      Centro
## Subgrupo      A      B      C      D
## QTL1 0.0000000 0.0000000 0.0000000 0.2538791
## QTL2 0.3220273 0.2512281 0.1345809 0.0000000
## SQTL 0.0000000 0.0000000 0.0382846 0.0000000

```

Table 1: Proporción datos por Subgrupo de peces en diferentes Centros

	A	B	C	D
QTL1	0.0000000	0.0000000	0.0000000	0.2538791
QTL2	0.3220273	0.2512281	0.1345809	0.0000000
SQTL	0.0000000	0.0000000	0.0382846	0.0000000

Los datos no están balanceados. Hay más observaciones para el Subgrupo QTL2 (70,7%) que para otros subgrupos. El Subgrupo en menor proporción es el SQTL que está presente solo en Centro C.