

QTL-SRS

Marcos Mancilla

16/10/2021

Realizando análisis exploratorio de datos

```
dat <- read_excel("dataset2.xls")
summary(dat)
```

```
##      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
```

```
head(dat)
```

```
## # 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>

str(dat)

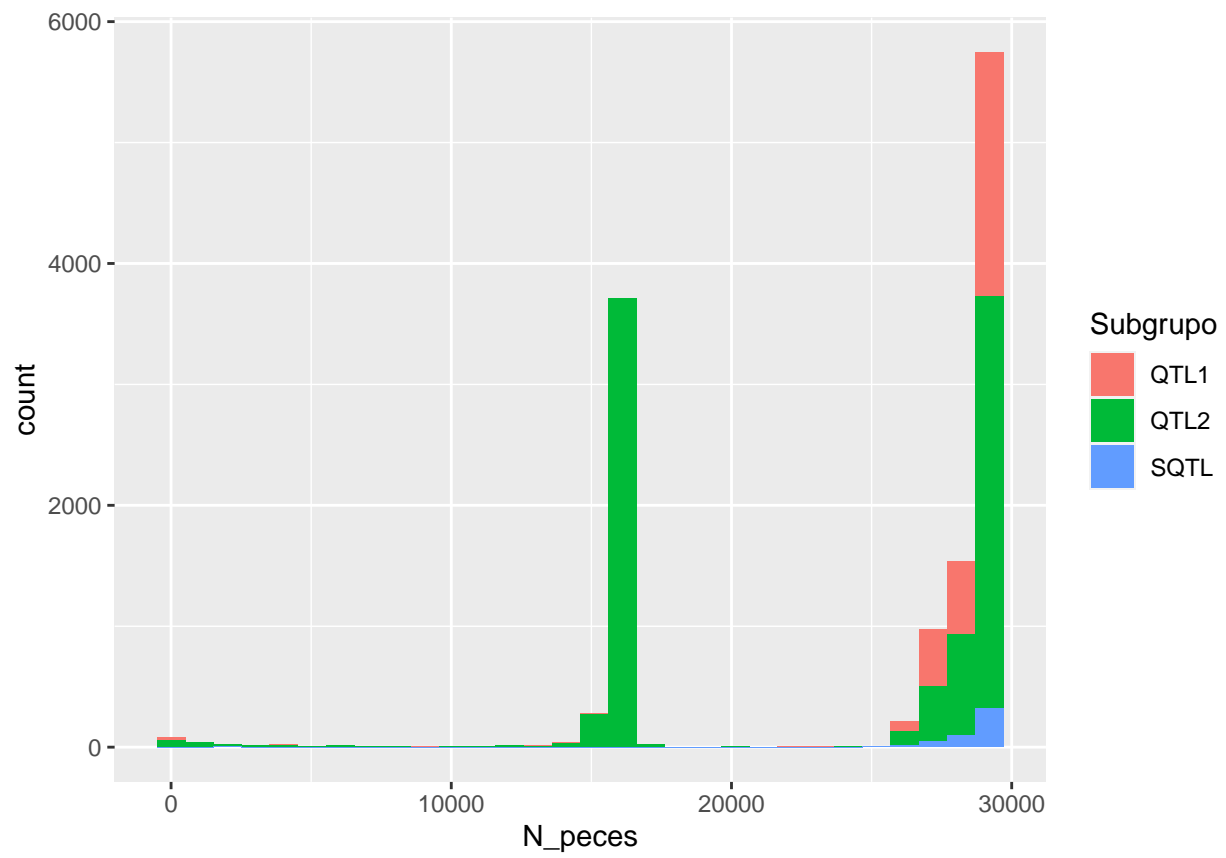
## tibble [12,825 x 12] (S3: tbl_df/tbl/data.frame)
##  $ Fecha          : POSIXct[1:12825], format: "2019-08-24" "2019-08-25" ...
##  $ Centro         : chr [1:12825] "A" "A" "A" "A" ...
##  $ Unidad         : chr [1:12825] "205" "205" "205" "205" ...
##  $ Subgrupo       : chr [1:12825] "QTL2" "QTL2" "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 ...
```

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

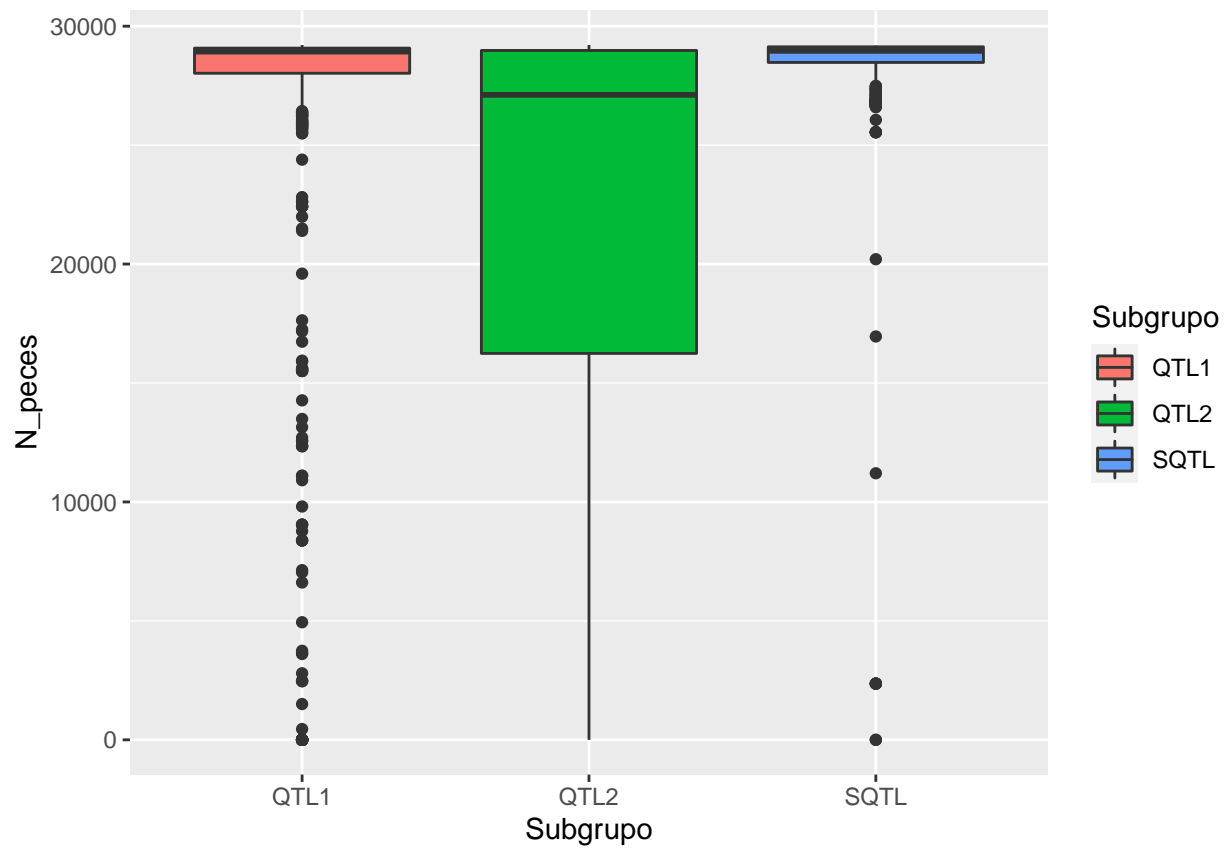
Peces por subgrupo

```
ggplot(dat, aes(x = N_peces, fill=Subgrupo)) +
  geom_histogram()

## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

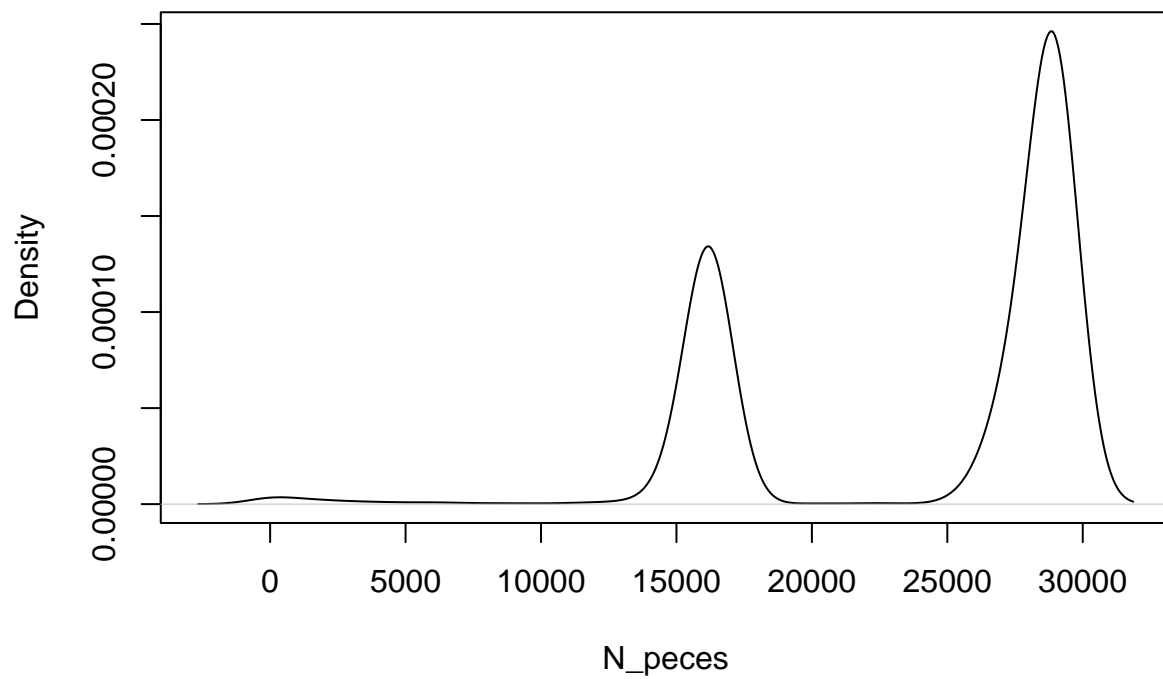


```
ggplot(dat, aes(x = Subgrupo, y = N_peces, fill=Subgrupo)) +
  geom_boxplot()
```



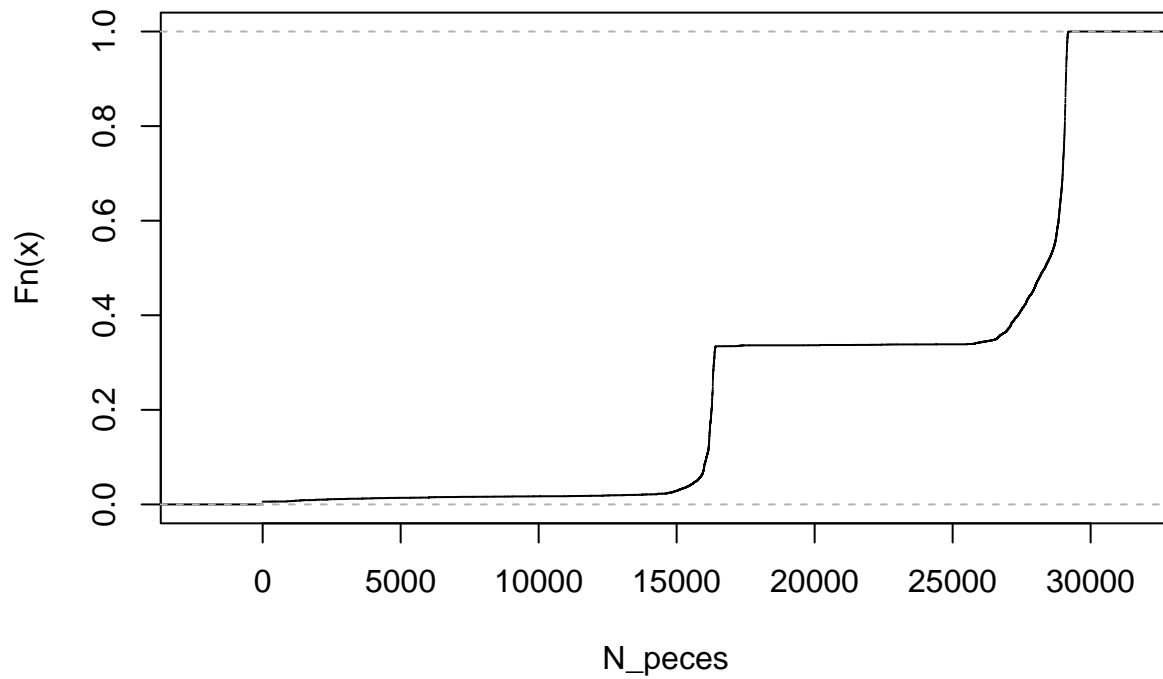
```
plot(density(dat$N_peces), main="Densidad empírica", xlab="N_peces")
```

Densidad empírica



```
plot(ecdf(dat$N_peces), main="Distribución acumulada empírica", xlab="N_peces")
```

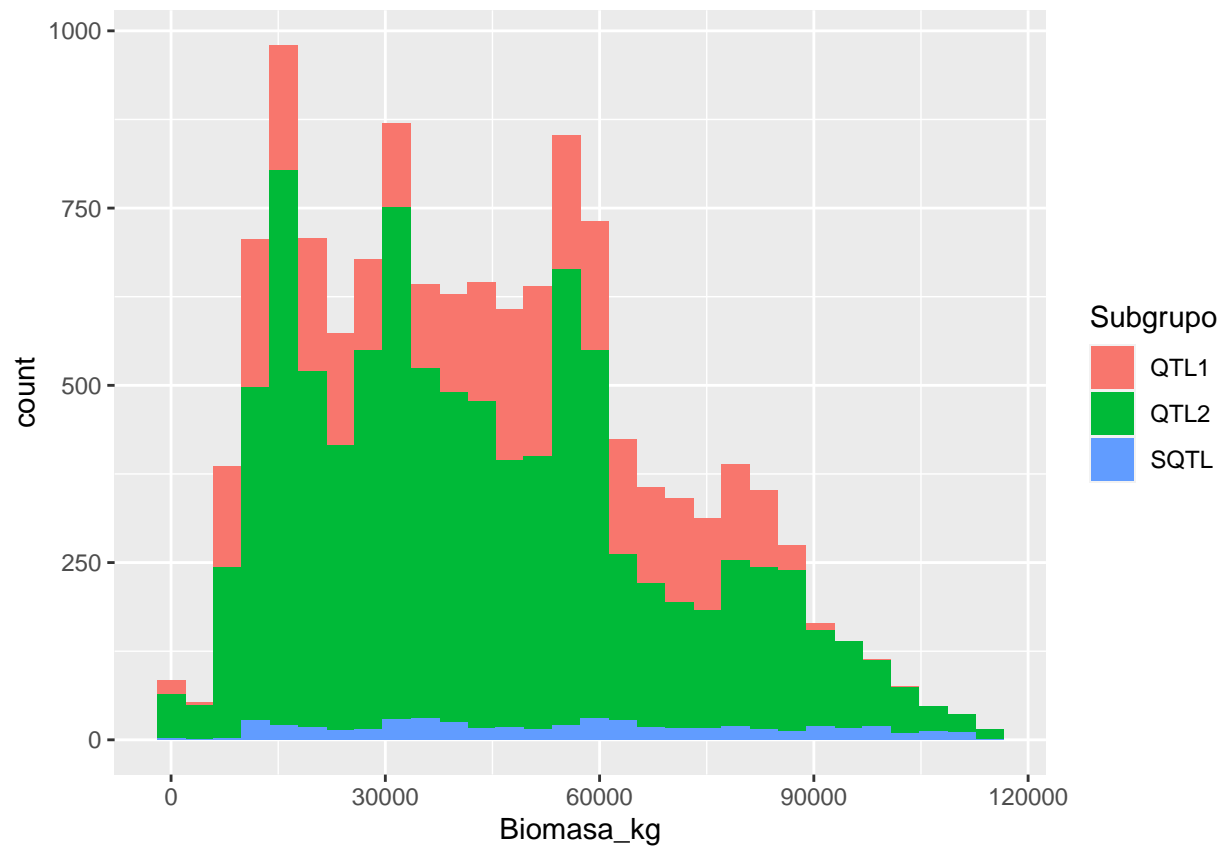
Distribución acumulada empírica



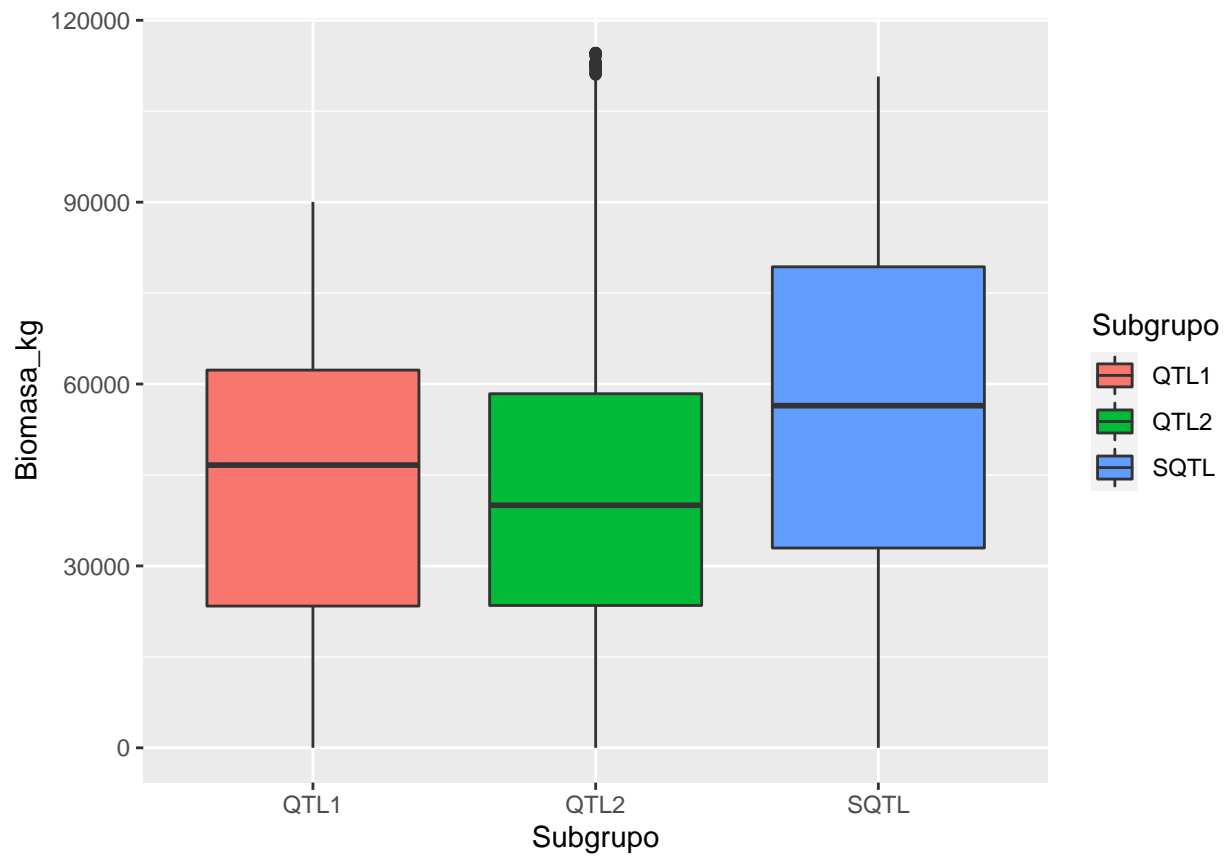
Biomasa viva por subgrupo

```
ggplot(dat, aes(x = Biomasa_kg, fill=Subgrupo)) +  
  geom_histogram()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

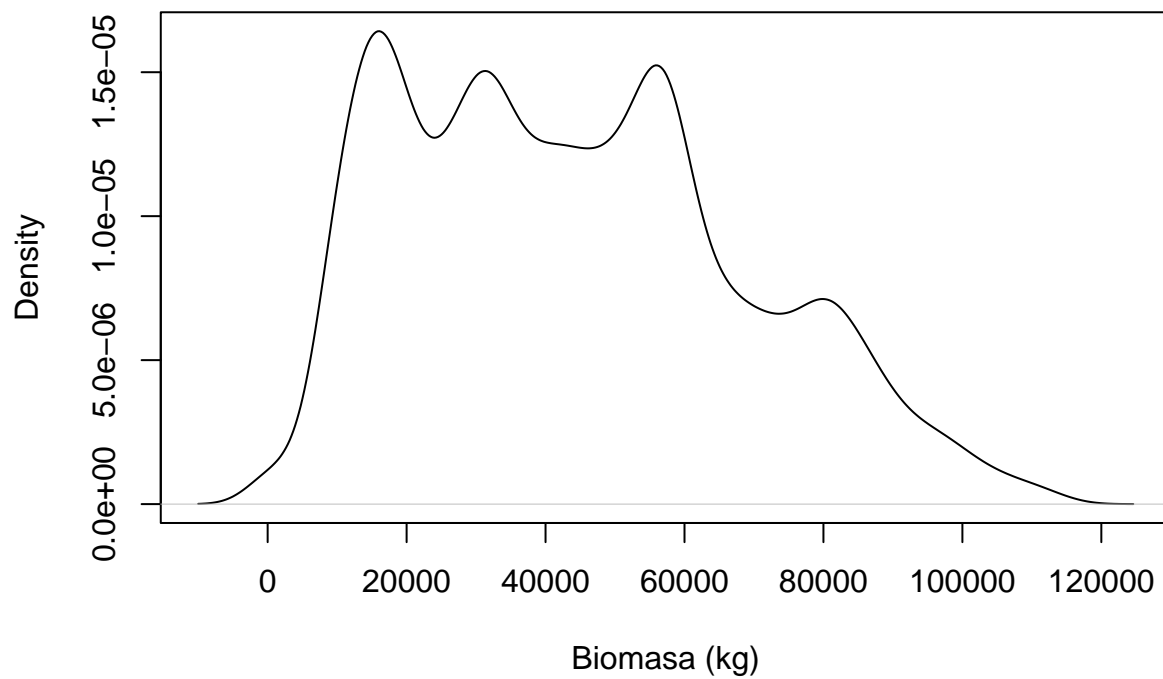


```
ggplot(dat, aes(x=Subgrupo, y = Biomasa_kg, fill=Subgrupo)) +
  geom_boxplot()
```



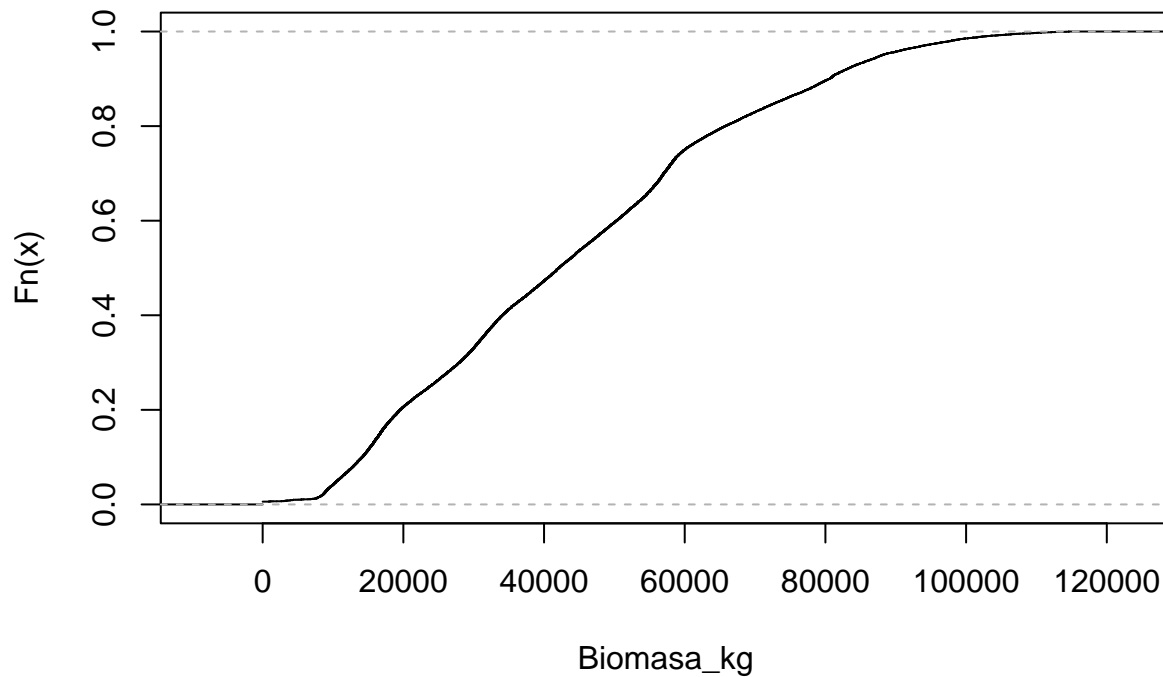
```
plot(density(dat$Biomasa_kg), main="Densidad empírica", xlab="Biomasa (kg)")
```

Densidad empírica



```
plot(ecdf(dat$Biomasa_kg), main="Distribución acumulada empírica", xlab="Biomasa_kg")
```

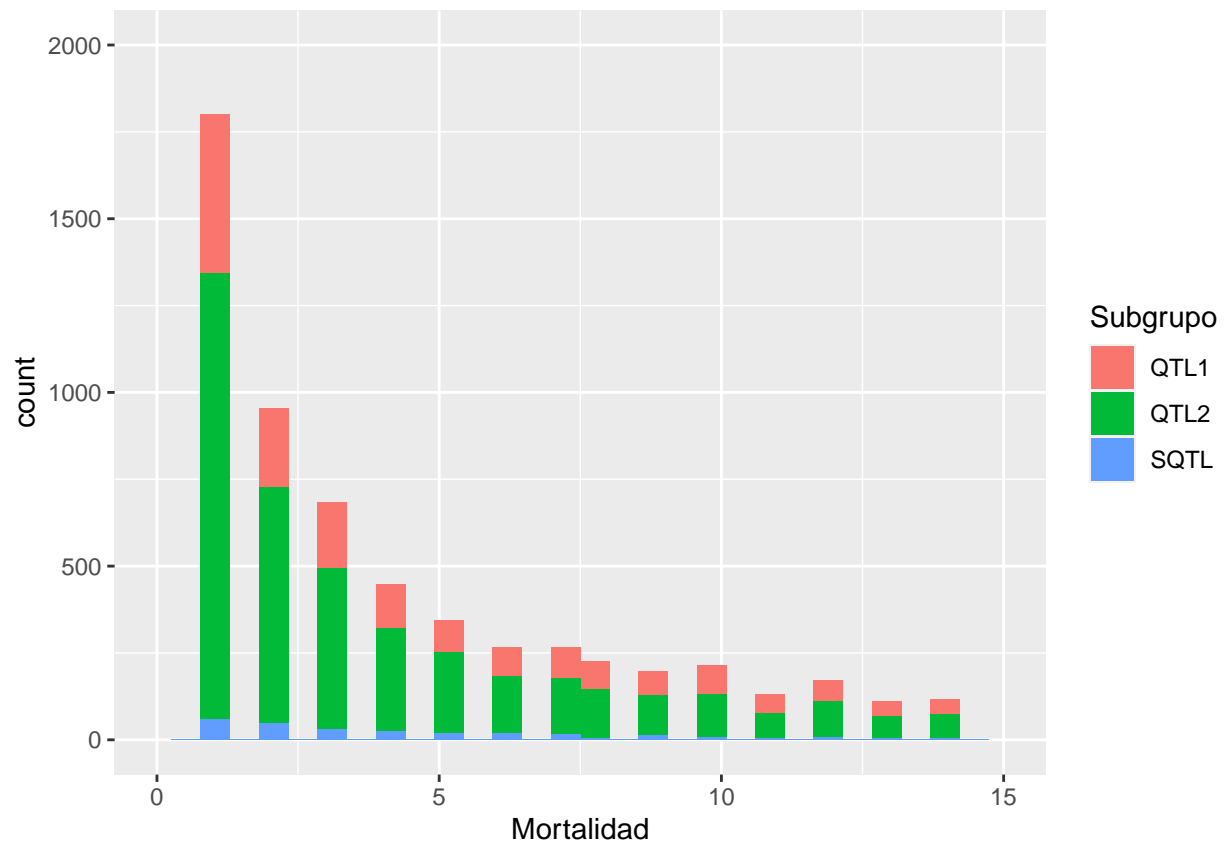
Distribución acumulada empírica



Mortalidad (n° peces) por subgrupo

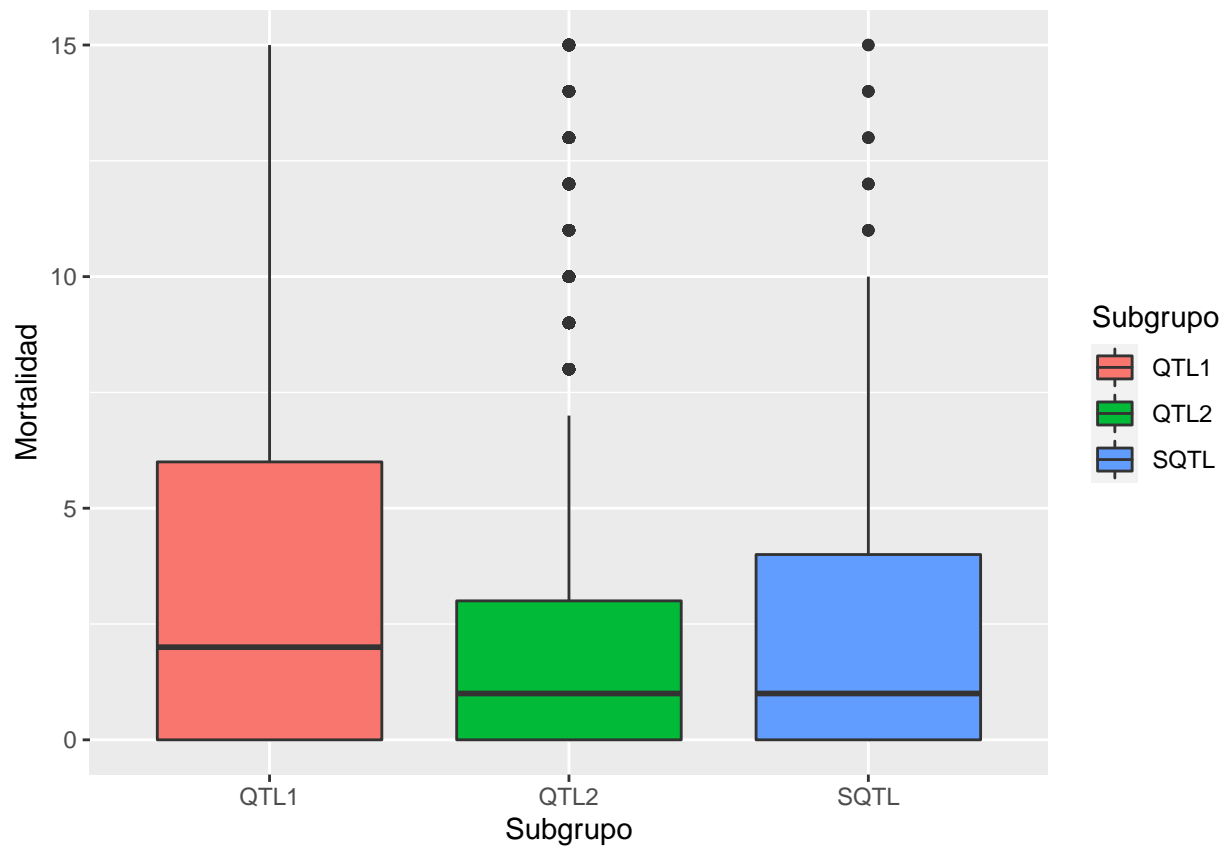
```
ggplot(dat, aes(x = Mortalidad, fill=Subgrupo)) +  
  geom_histogram()+ scale_x_continuous(limit = c(0,15))+ scale_y_continuous(limit = c(0,2000))
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## Warning: Removed 2139 rows containing non-finite values (stat_bin).  
## Warning: Removed 6 rows containing missing values (geom_bar).
```

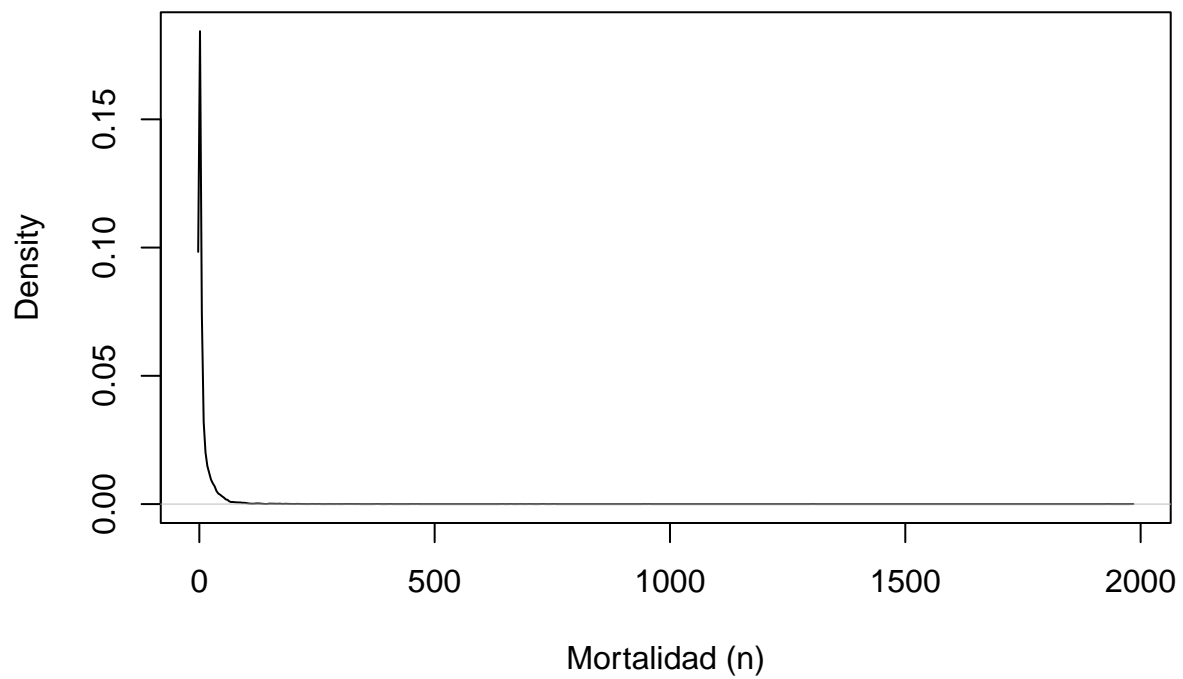
```
ggplot(dat, aes(x=Subgrupo, y = Mortalidad, fill=Subgrupo)) +
  geom_boxplot()+ scale_y_continuous(limit = c(0,15))
```

```
## Warning: Removed 2139 rows containing non-finite values (stat_boxplot).
```



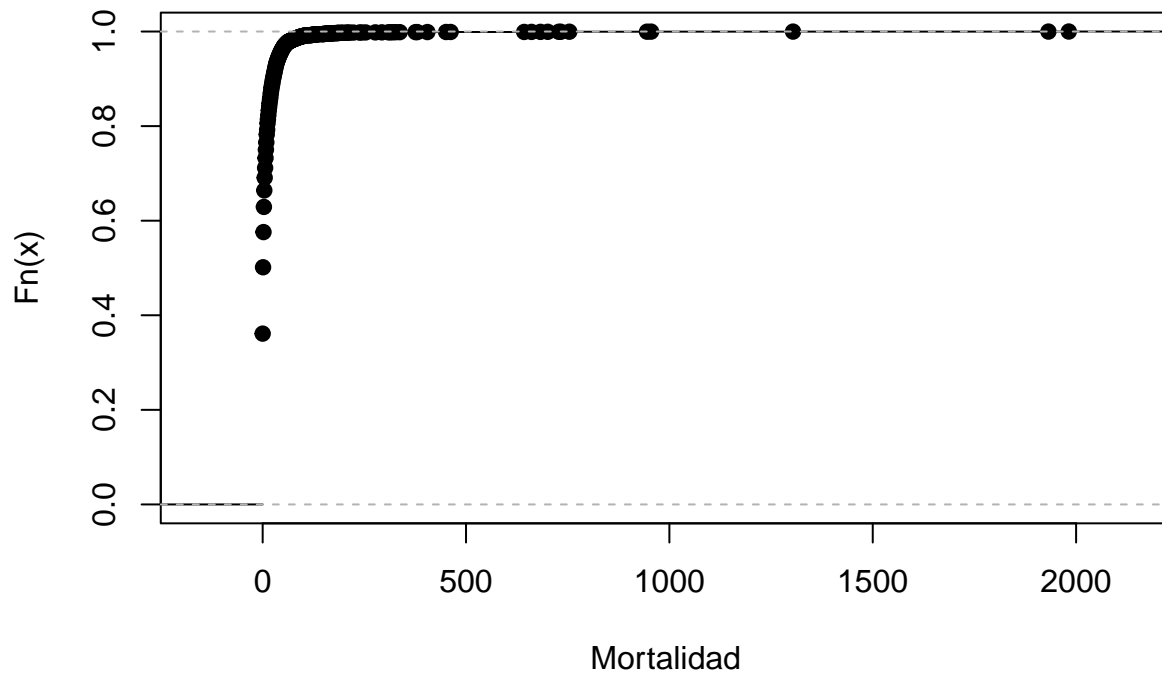
```
plot(density(dat$Mortalidad), main="Densidad empírica", xlab="Mortalidad (n)")
```

Densidad empírica



```
plot(ecdf(dat$Mortalidad), main="Distribución acumulada empírica", xlab="Mortalidad")
```

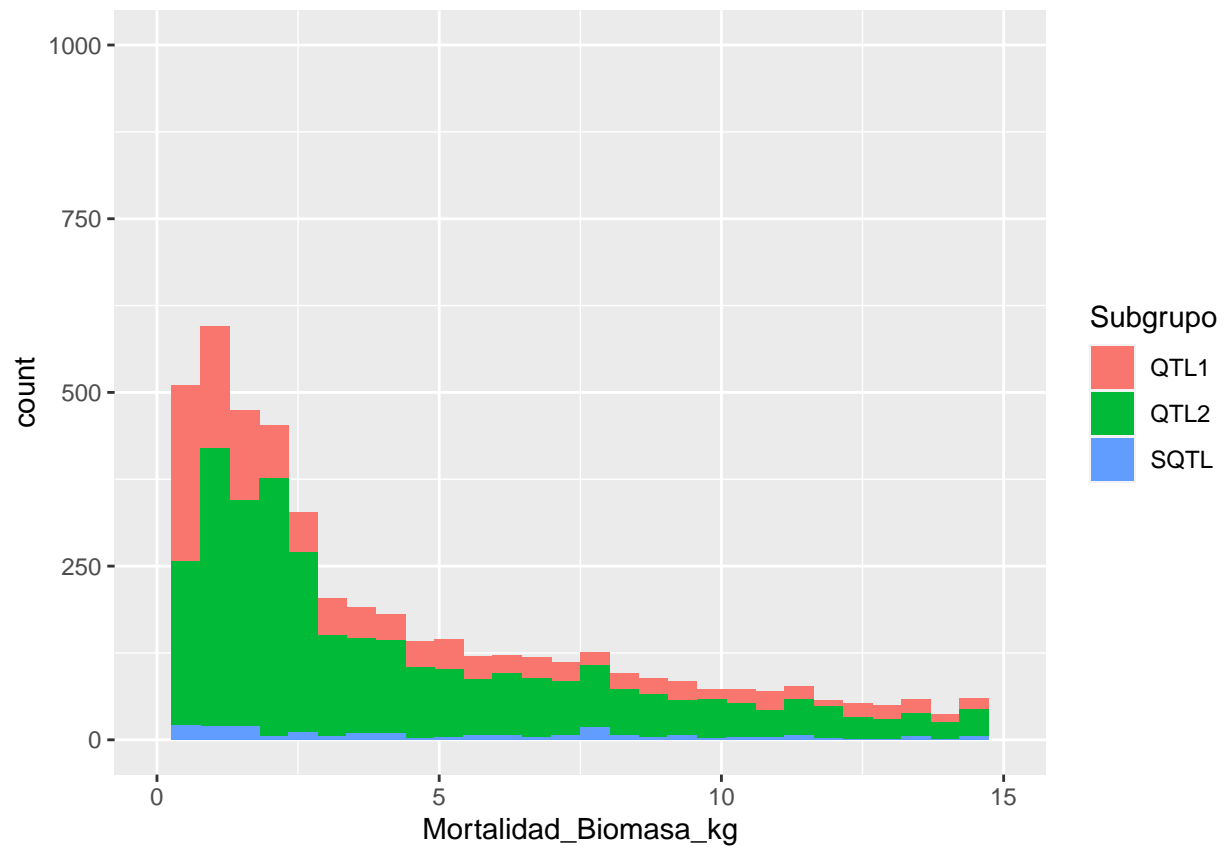
Distribución acumulada empírica



Biomasa de la mortalidad por subgrupo

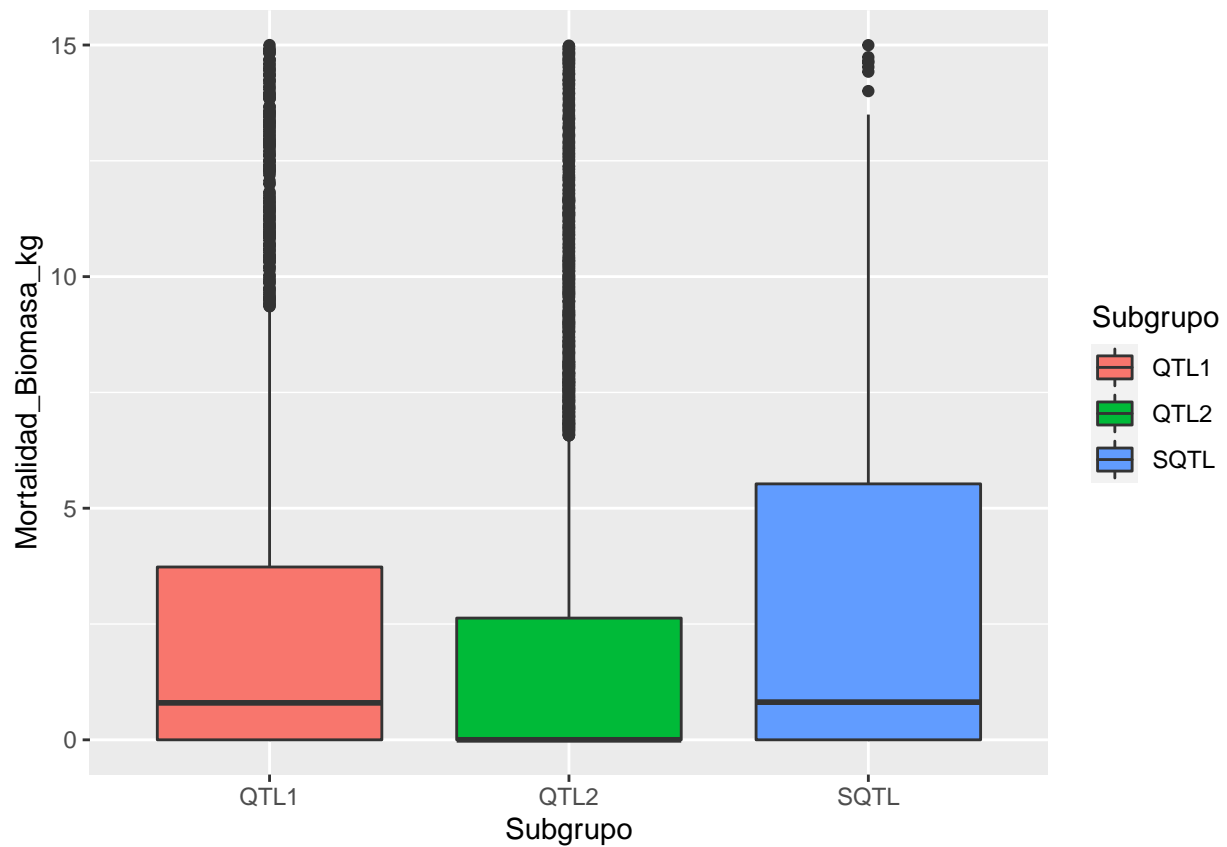
```
ggplot(dat, aes(x = Mortalidad_Biomasa_kg, fill=Subgrupo)) +  
  geom_histogram()+ scale_x_continuous(limit = c(0,15))+ scale_y_continuous(limit = c(0,1000))
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## Warning: Removed 3470 rows containing non-finite values (stat_bin).  
## Warning: Removed 6 rows containing missing values (geom_bar).
```



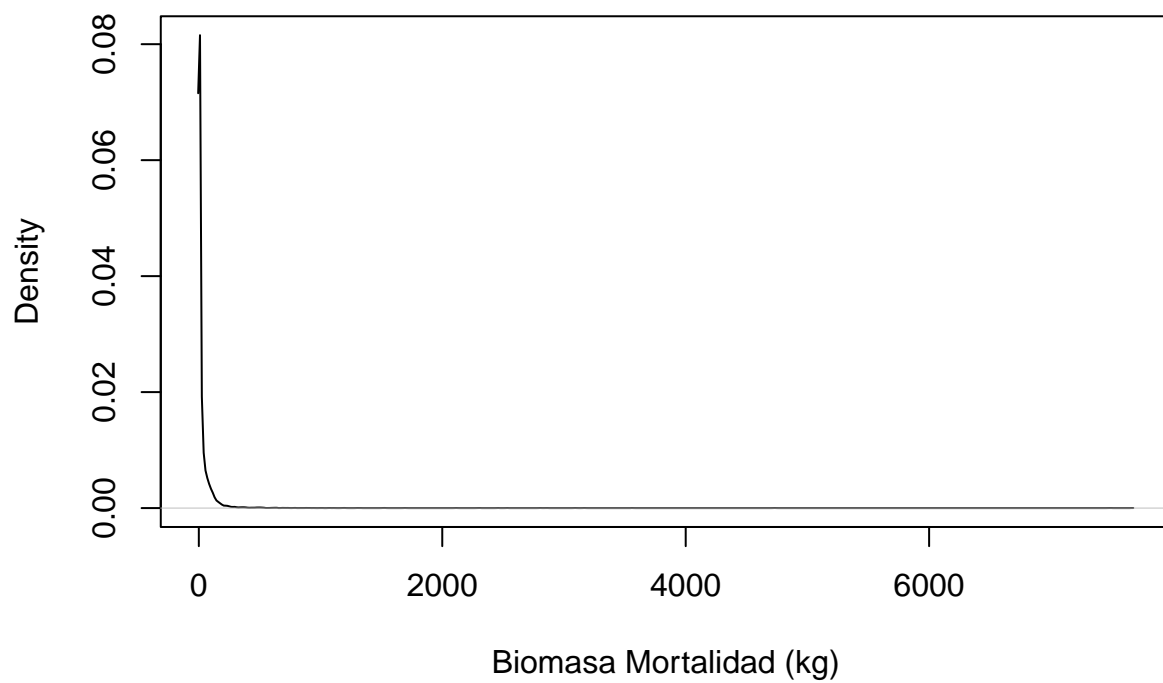
```
ggplot(dat, aes(x=Subgrupo, y = Mortalidad_Biomasa_kg, fill=Subgrupo)) +
  geom_boxplot()+ scale_y_continuous(limit = c(0,15))
```

Warning: Removed 3470 rows containing non-finite values (stat_boxplot).

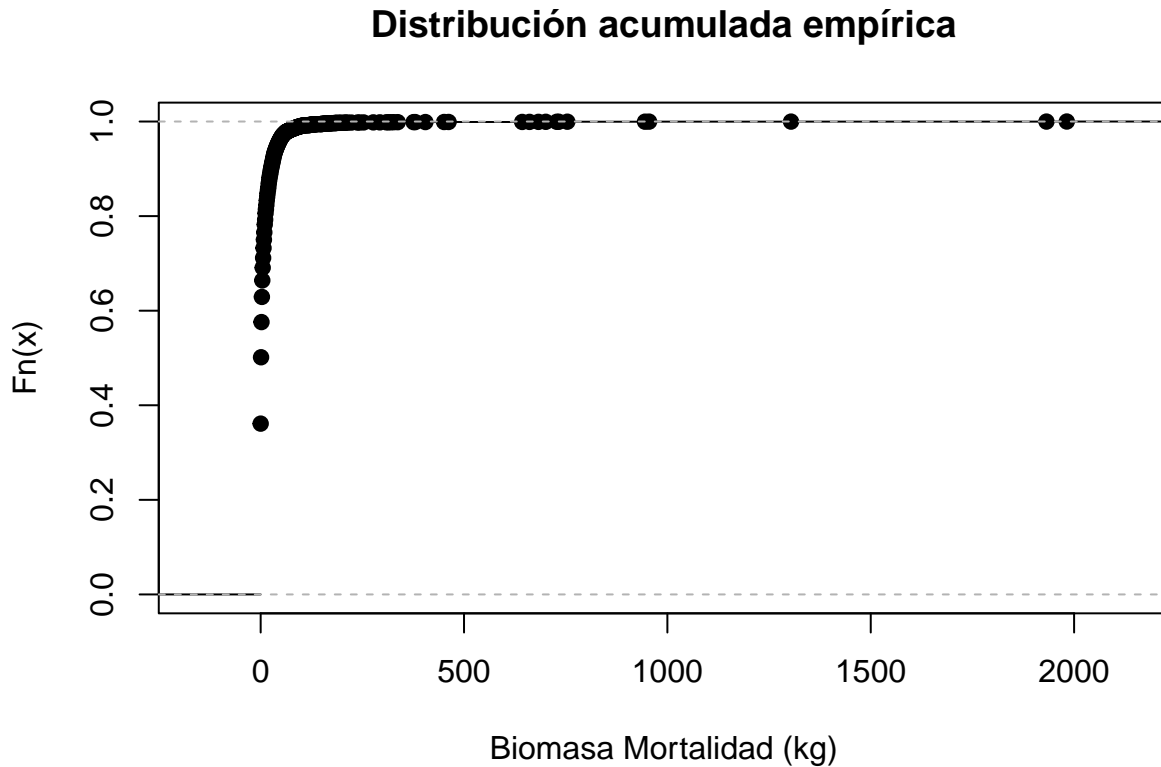


```
plot(density(dat$Mortalidad_Biomasa_kg), main="Densidad empírica", xlab="Biomasa Mortalidad (kg)")
```

Densidad empírica



```
plot(ecdf(dat$Mortalidad), main="Distribución acumulada empírica", xlab="Biomasa Mortalidad (kg)")
```



```
dat1 <- dat
table(dat1$Subgrupo)
```

```
##
## QTL1 QTL2 SCTL
## 3256 9078 491
```

```
table(dat1$Centro)
```

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

```
dat1$Centro <- as.factor(dat1$Centro)
dat1$Subgrupo <- as.factor(dat1$Subgrupo)
dat1$Unidad <- as.factor(dat1$Unidad)
str(dat1)
```

```
## 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 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 ...
```

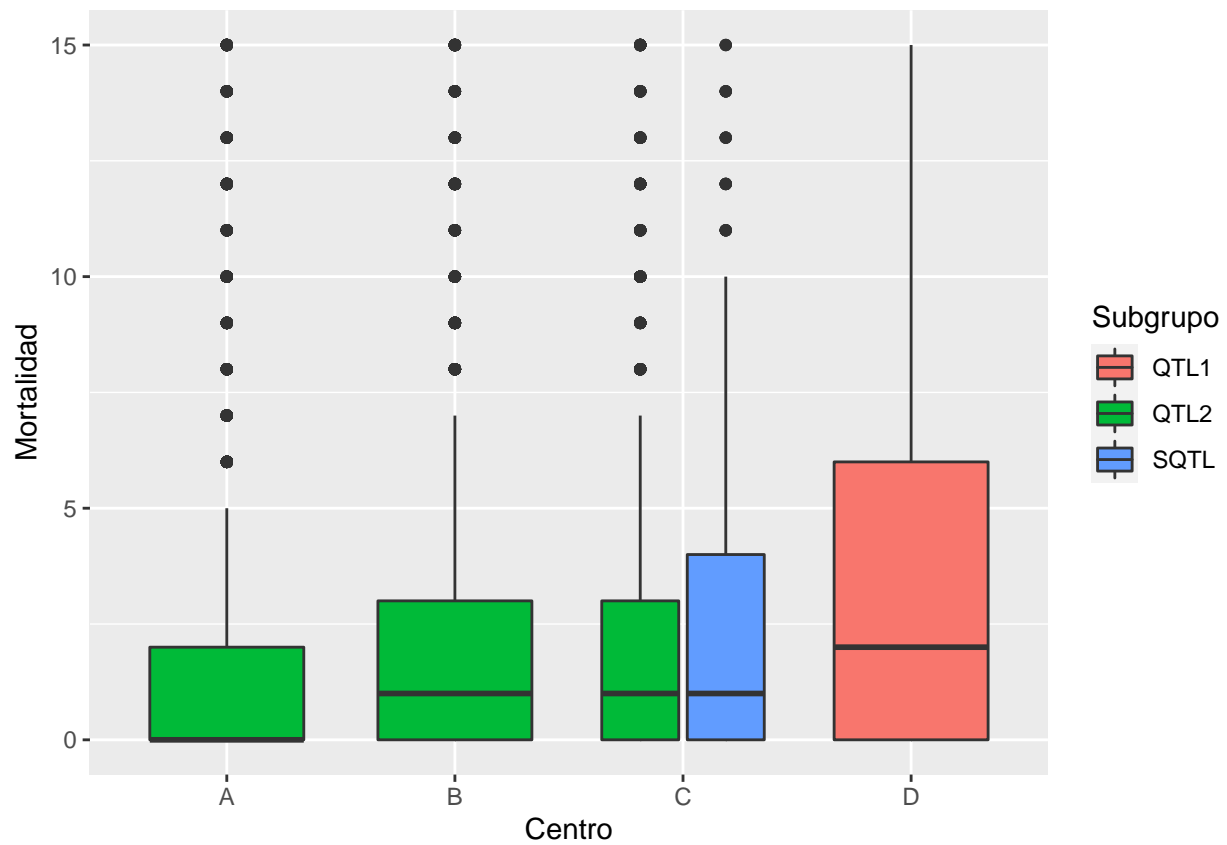
```
summary(dat1)
```

```
##      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
##
```

Hacemos boxplot de variable respuesta en función de otras variables

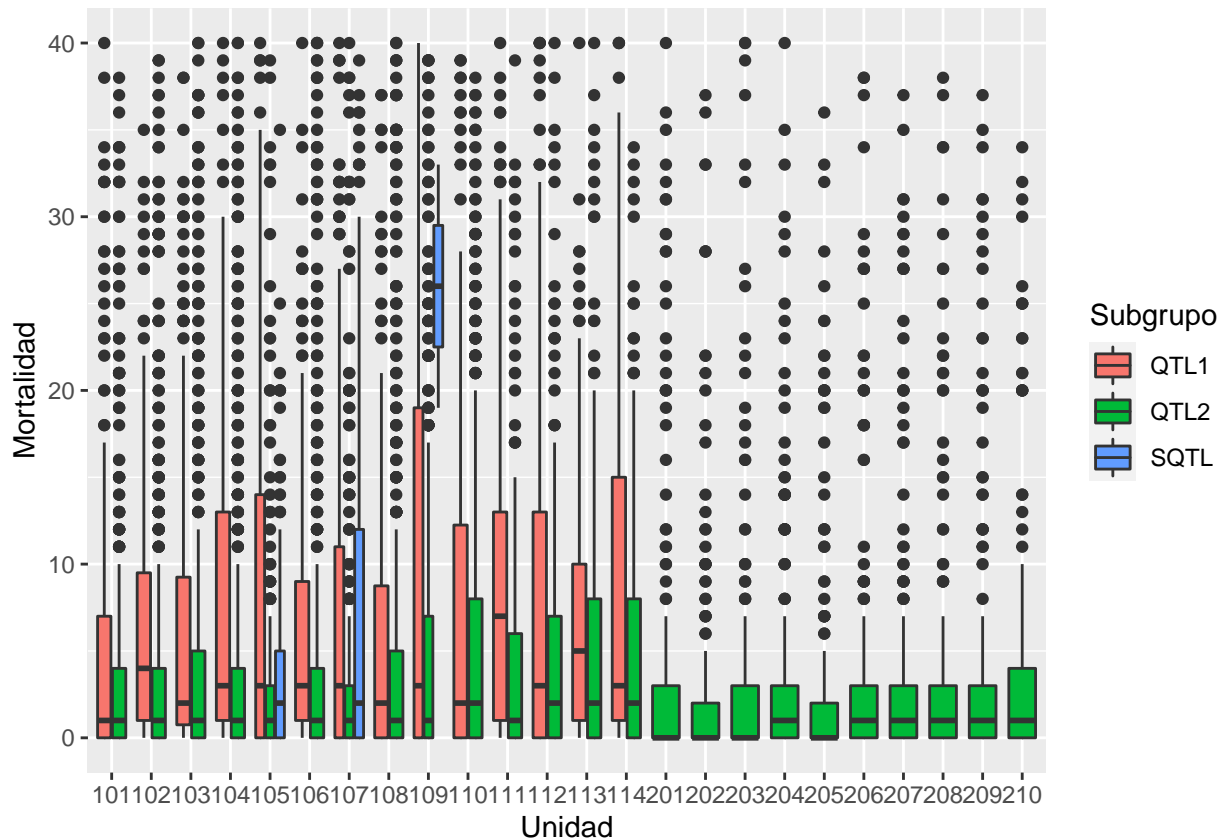
```
ggplot(dat1, aes(x=Centro, y = Mortalidad, fill=Subgrupo)) +
  geom_boxplot()+ scale_y_continuous(limit = c(0,15))
```

```
## Warning: Removed 2139 rows containing non-finite values (stat_boxplot).
```



```
ggplot(dat1, aes(x=Unidad, y = Mortalidad, fill=Subgrupo)) +
  geom_boxplot()+ scale_y_continuous(limit = c(0,40))
```

```
## Warning: Removed 670 rows containing non-finite values (stat_boxplot).
```

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

```
table(dat$Subgrupo, dat$Centro)
```

```
##
##           A      B      C      D
##  QTL1      0      0      0 3256
##  QTL2 4130 3222 1726      0
##  SCTL      0      0    491      0
```

```
tabla1 <- with(dat, table(Subgrupo, Centro))
prop.table(tabla1)
```

```
##           Centro
## Subgrupo      A      B      C      D
##   QTL1 0.0000000 0.0000000 0.0000000 0.2538791
##   QTL2 0.3220273 0.2512281 0.1345809 0.0000000
##   SCTL 0.0000000 0.0000000 0.0382846 0.0000000
```

```
tabla1_prop <- with(dat, prop.table(tabla1))
```

```
knitr::kable(tabla1_prop, caption = "Proporción datos por Subgrupo de peces en diferentes Centros")
```

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.

```
df1 <- dat %>%
  select(Subgrupo, Centro, Unidad, Biomasa_kg, N_peces) %>%
  group_by(Subgrupo, Centro, Unidad) %>%
  summarize(Maximo = max(N_peces), Minimo = min(N_peces), Max_Biom = round(max(Biomasa_kg), 0))
```

`summarise()` has grouped output by 'Subgrupo', 'Centro'. You can override using the `.groups` argument

```
knitr::kable(df1, caption = "N° de peces por unidad (jaula)")
```

Table 2: N° de peces por unidad (jaula)

Subgrupo	Centro	Unidad	Maximo	Minimo	Max_Biom
QTL1	D	101	29132	0	79965
QTL1	D	102	29093	0	86955
QTL1	D	103	29209	0	85370
QTL1	D	104	29146	0	82155
QTL1	D	105	29209	0	83898
QTL1	D	106	29116	0	90041
QTL1	D	107	29125	0	82005
QTL1	D	108	29144	0	82906
QTL1	D	109	29179	0	81748
QTL1	D	110	29176	0	81228
QTL1	D	111	29209	0	85700
QTL1	D	112	29160	0	86777
QTL1	D	113	29209	0	87577
QTL1	D	114	29048	0	81422
QTL2	A	101	16259	0	58239
QTL2	A	102	16393	0	64257
QTL2	A	103	16327	0	57385
QTL2	A	104	16430	0	60023
QTL2	A	105	16357	0	56487
QTL2	A	106	16430	0	59141
QTL2	A	107	16363	0	58682
QTL2	A	108	16430	0	59910
QTL2	A	109	16340	0	57885
QTL2	A	110	16379	0	58648
QTL2	A	201	16402	0	59676
QTL2	A	202	16416	0	58887

Subgrupo	Centro	Unidad	Maximo	Minimo	Max_Biom
QTL2	A	203	16393	0	57330
QTL2	A	204	16413	0	58648
QTL2	A	205	16416	0	56185
QTL2	A	206	16394	0	57831
QTL2	A	207	16430	0	57951
QTL2	A	208	16418	0	57117
QTL2	A	209	16430	0	58669
QTL2	A	210	16409	0	56785
QTL2	B	101	29164	0	109665
QTL2	B	102	28956	26947	107660
QTL2	B	103	29209	0	101358
QTL2	B	104	29085	0	109601
QTL2	B	105	29209	26606	105570
QTL2	B	106	29125	27077	107523
QTL2	B	107	29189	26952	100875
QTL2	B	108	29118	0	108229
QTL2	B	109	29196	0	84264
QTL2	B	110	29209	27633	111307
QTL2	B	111	29186	0	105306
QTL2	B	112	29174	0	105503
QTL2	B	113	29196	27896	114599
QTL2	B	114	29209	27941	114073
QTL2	C	101	29171	0	99127
QTL2	C	102	29188	0	93891
QTL2	C	103	29186	0	96741
QTL2	C	104	29172	0	87664
QTL2	C	105	4827	0	2011
QTL2	C	106	29130	0	92344
QTL2	C	108	29124	0	84824
QTL2	C	109	29137	25786	97986
QTL2	C	110	28900	0	88063
SQTL	C	105	29184	2362	110722
SQTL	C	107	29185	0	101705
SQTL	C	109	16956	0	5387

El análisis anterior indica que la mayoría de las unidades (jaulas) se cosecharon al final del periodo de observación (Mínimo = 0), quedando solo 9 jaulas con peces. Si sumamos el número de peces por Subgrupo, confirmamos que el Subgrupo QTL2 está sobrerrepresentado. Aprovechamos de calcular la biomasa que llegaron a desarrollar estos Subgrupos.

```
df2 <-df1 %>%
  select(Subgrupo, Maximo, Max_Biom)%>%
  group_by(Subgrupo)%>%
  summarize(N_Peces=sum(Maximo), Biomasa_total=sum(Max_Biom))
```

```
## Adding missing grouping variables: `Centro`
```

```
knitr::kable(df2,caption = "N° de peces y biomasa por Subgrupo")
```

Table 3: N° de peces y biomasa por Subgrupo

Subgrupo	N_Peces	Biomasa_total
QTL1	408155	1177747
QTL2	973889	3397920
SQTL	75325	217814

Calculamos ahora la mortalidad asociada a cada Unidad.

```
df3 <-dat %>%
  select(Subgrupo, Centro, Unidad, Mortalidad, Mortalidad_Biomasa_kg)%>%
  group_by(Subgrupo, Centro, Unidad)%>%
  summarize(Mort_peces = sum(Mortalidad), Biomas_mort=round(sum(Mortalidad_Biomasa_kg), 0))
```

`summarise()` has grouped output by 'Subgrupo', 'Centro'. You can override using the `.groups` argument

```
knitr::kable(df3,caption = "Mortalidad de peces por Unidad")
```

Table 4: Mortalidad de peces por Unidad

Subgrupo	Centro	Unidad	Mort_peces	Biomas_mort
QTL1	D	101	2254	5592
QTL1	D	102	1928	3705
QTL1	D	103	2071	5278
QTL1	D	104	2321	5398
QTL1	D	105	2770	6750
QTL1	D	106	2015	4019
QTL1	D	107	2908	7157
QTL1	D	108	1607	3683
QTL1	D	109	3753	8638
QTL1	D	110	2708	6594
QTL1	D	111	3642	8430
QTL1	D	112	2754	6372
QTL1	D	113	2507	5239
QTL1	D	114	2983	6810
QTL2	A	101	2627	8556
QTL2	A	102	1677	5642
QTL2	A	103	829	2250
QTL2	A	104	2413	8603
QTL2	A	105	2126	6893
QTL2	A	106	2179	7737
QTL2	A	107	934	2764
QTL2	A	108	1225	3777
QTL2	A	109	2842	9722
QTL2	A	110	1517	4174
QTL2	A	201	1628	5678
QTL2	A	202	1881	6535
QTL2	A	203	2298	8033
QTL2	A	204	1922	6531
QTL2	A	205	2815	9903

Subgrupo	Centro	Unidad	Mort_peces	Biomasa_mort
QTL2	A	206	1828	6238
QTL2	A	207	1841	6332
QTL2	A	208	2287	7744
QTL2	A	209	1770	6056
QTL2	A	210	2160	7120
QTL2	B	101	1317	3188
QTL2	B	102	2262	5962
QTL2	B	103	2005	4898
QTL2	B	104	2143	6061
QTL2	B	105	2603	7418
QTL2	B	106	2132	5365
QTL2	B	107	2257	5867
QTL2	B	108	2133	5746
QTL2	B	109	2949	8246
QTL2	B	110	1576	4124
QTL2	B	111	1231	2925
QTL2	B	112	1551	3674
QTL2	B	113	1313	3187
QTL2	B	114	1268	2888
QTL2	C	101	2233	7203
QTL2	C	102	1802	5075
QTL2	C	103	2894	9729
QTL2	C	104	2412	6507
QTL2	C	105	4	2
QTL2	C	106	2114	6030
QTL2	C	108	3260	9086
QTL2	C	109	3371	8675
QTL2	C	110	3513	8699
SQTL	C	105	849	2134
SQTL	C	107	2564	6660
SQTL	C	109	52	17

Ahora, calculamos la mortalidad asociada por Subgrupo.

```
df4 <-df3 %>%
  select(Subgrupo, Mort_peces, Biomasa_mort)%>%
  group_by(Subgrupo)%>%
  summarize(Mortalidad = sum(Mort_peces), Biomasa_dead = sum(Biomasa_mort))

## Adding missing grouping variables: `Centro`
knitr::kable(df4,caption = "Mortalidad de peces por Subgrupo")
```

Table 5: Mortalidad de peces por Subgrupo

Subgrupo	Mortalidad	Biomasa_dead
QTL1	36221	83665
QTL2	87142	260843
SQTL	3465	8811

Queremos agregar a esta tabla las columnas de n° de peces y biomasa total del df2.

```
Tabla3 <- merge(df2,df4,by="Subgrupo")
knitr::kable(Tabla3,caption = "Resumen datos de Subgrupos")
```

Table 6: Resumen datos de Subgrupos

Subgrupo	N_Peces	Biomasa_total	Mortalidad	Biomasa_dead
QTL1	408155	1177747	36221	83665
QTL2	973889	3397920	87142	260843
SQTL	75325	217814	3465	8811