

QTL-SRS

Marcos Mancilla

27 November 2021

Realizando análisis exploratorio de datos

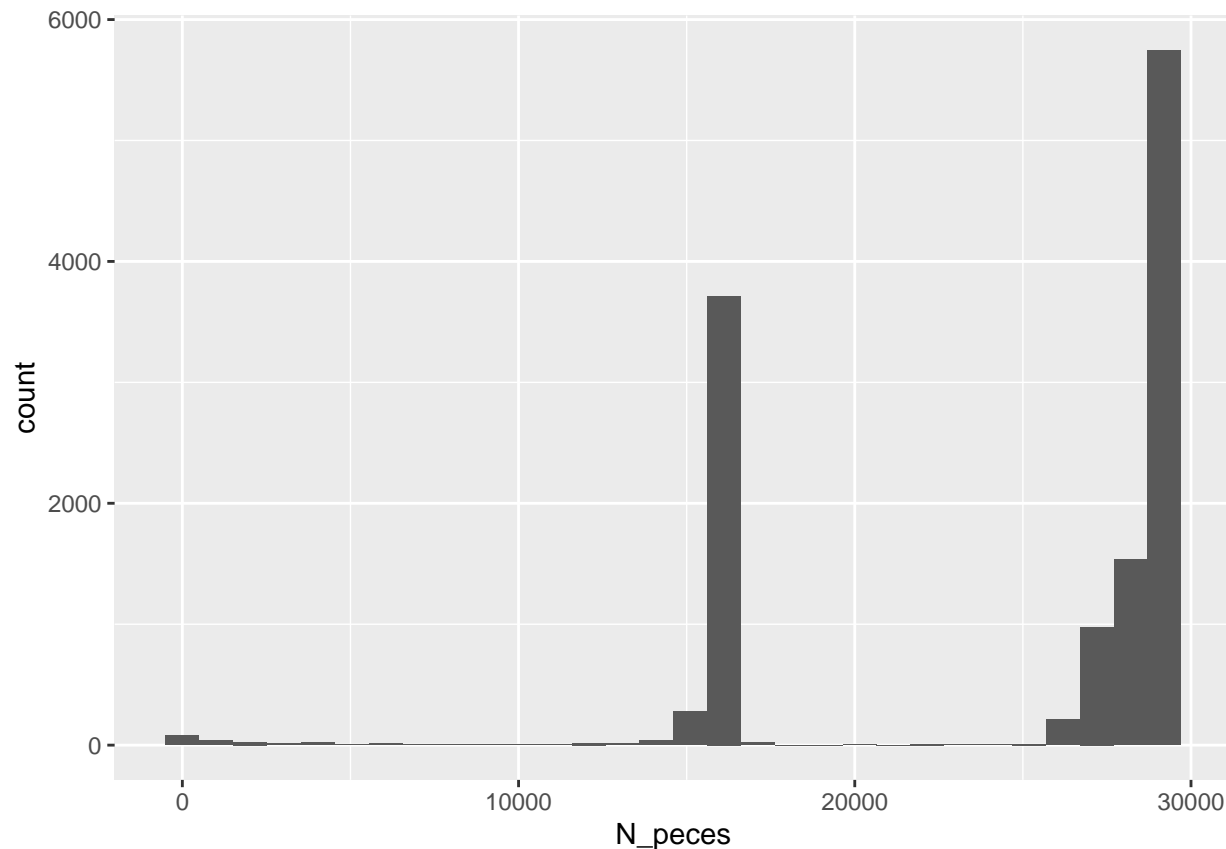
```
##      Fecha                Centro                Unidad
##  Min.   :2019-02-09 00:00:00  Length:12825      Length:12825
##  1st Qu.:2019-06-01 00:00:00  Class :character  Class :character
##  Median :2019-07-26 00:00:00  Mode  :character  Mode  :character
##  Mean   :2019-07-25 16:53:46
##  3rd 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>
```

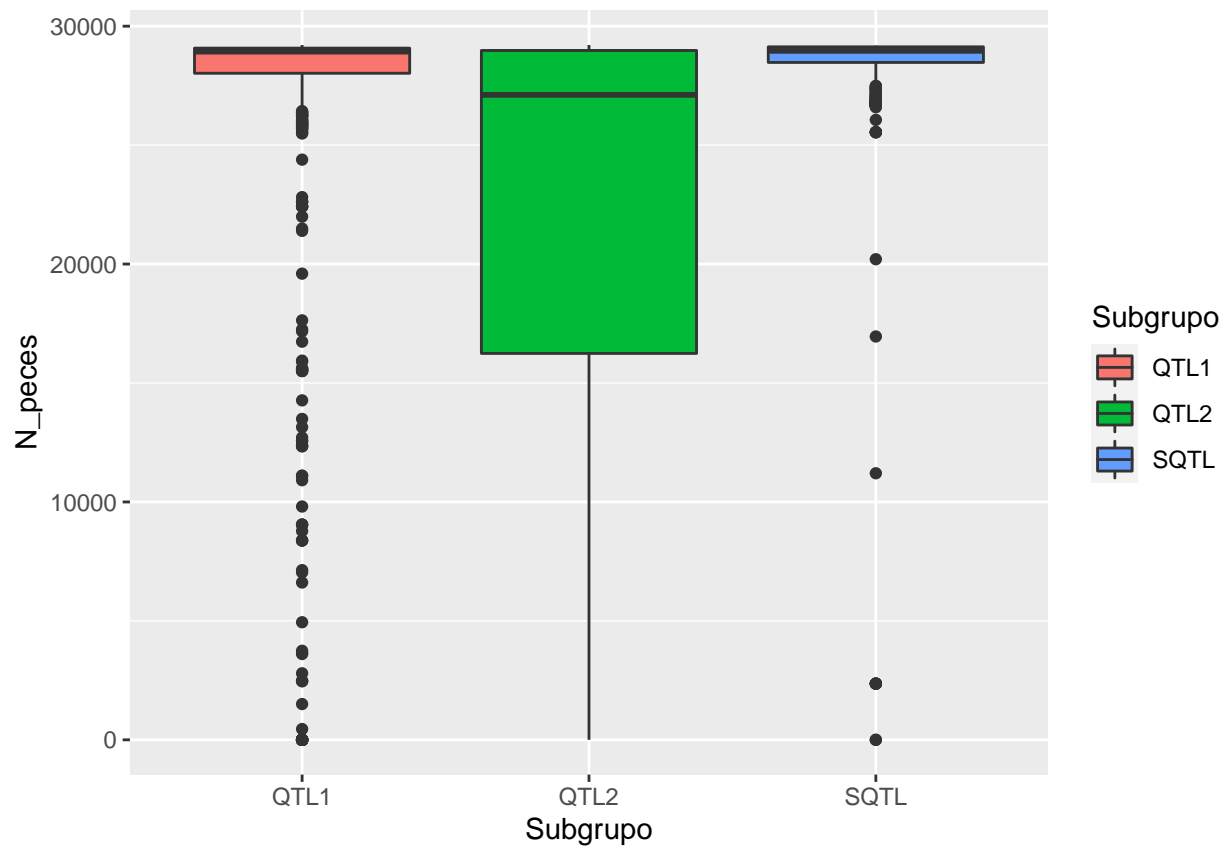
```
## tibble [12,825 x 12] (S3: tbl_df/tbl/data.frame)
## $ Fecha          : POSIXct[1:12825], format: "2019-08-24" "2019-08-25" ...
## $ Centro         : chr [1:12825] "A" "A" "A" "A" ...
## $ Unidad         : chr [1:12825] "205" "205" "205" "205" ...
## $ Subgrupo       : chr [1:12825] "QTL2" "QTL2"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

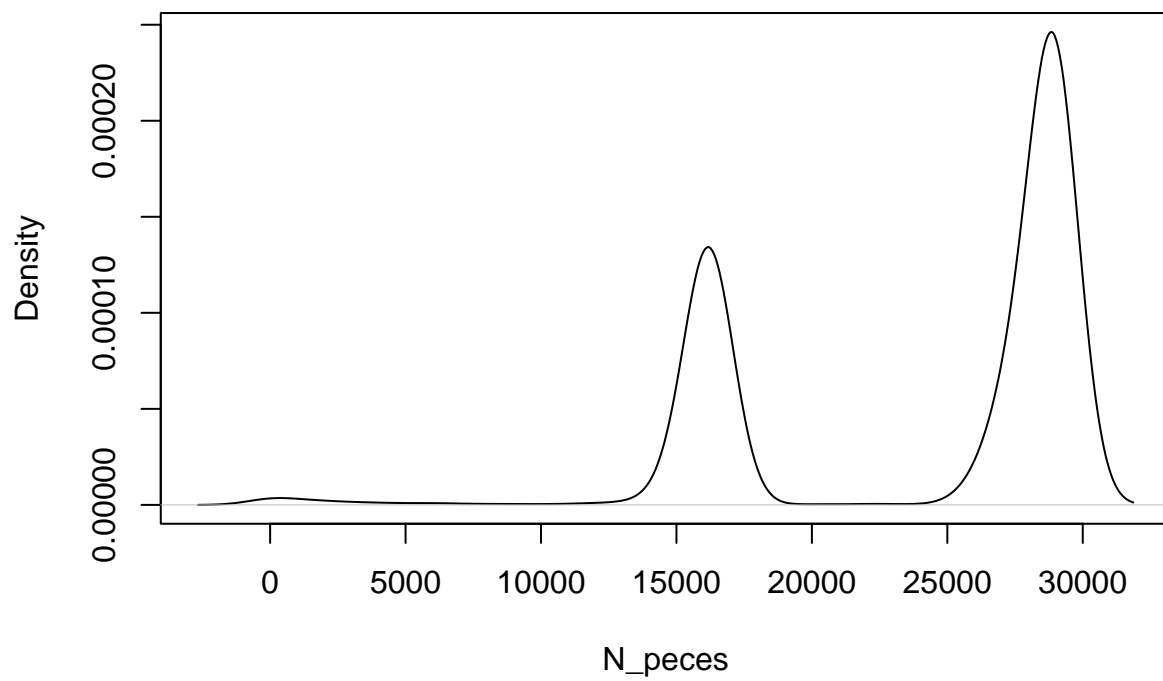
Graficamos variables de interés (histogramas, boxplots, densidad y densidad acumulada).

Peces por subgrupo

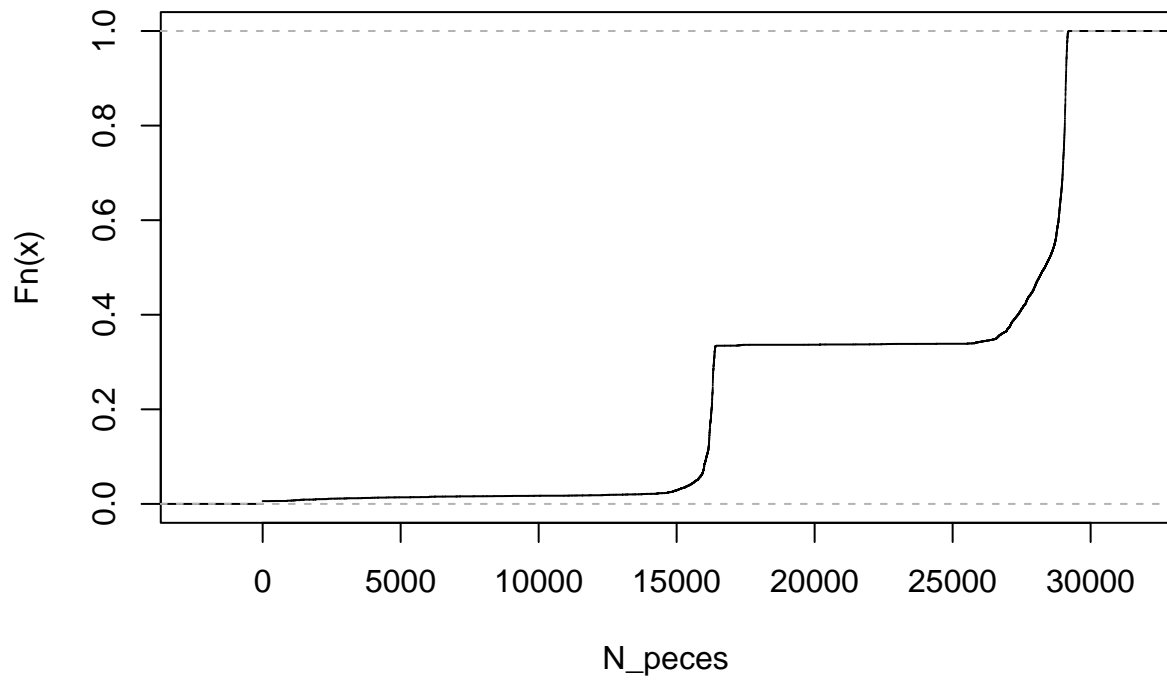




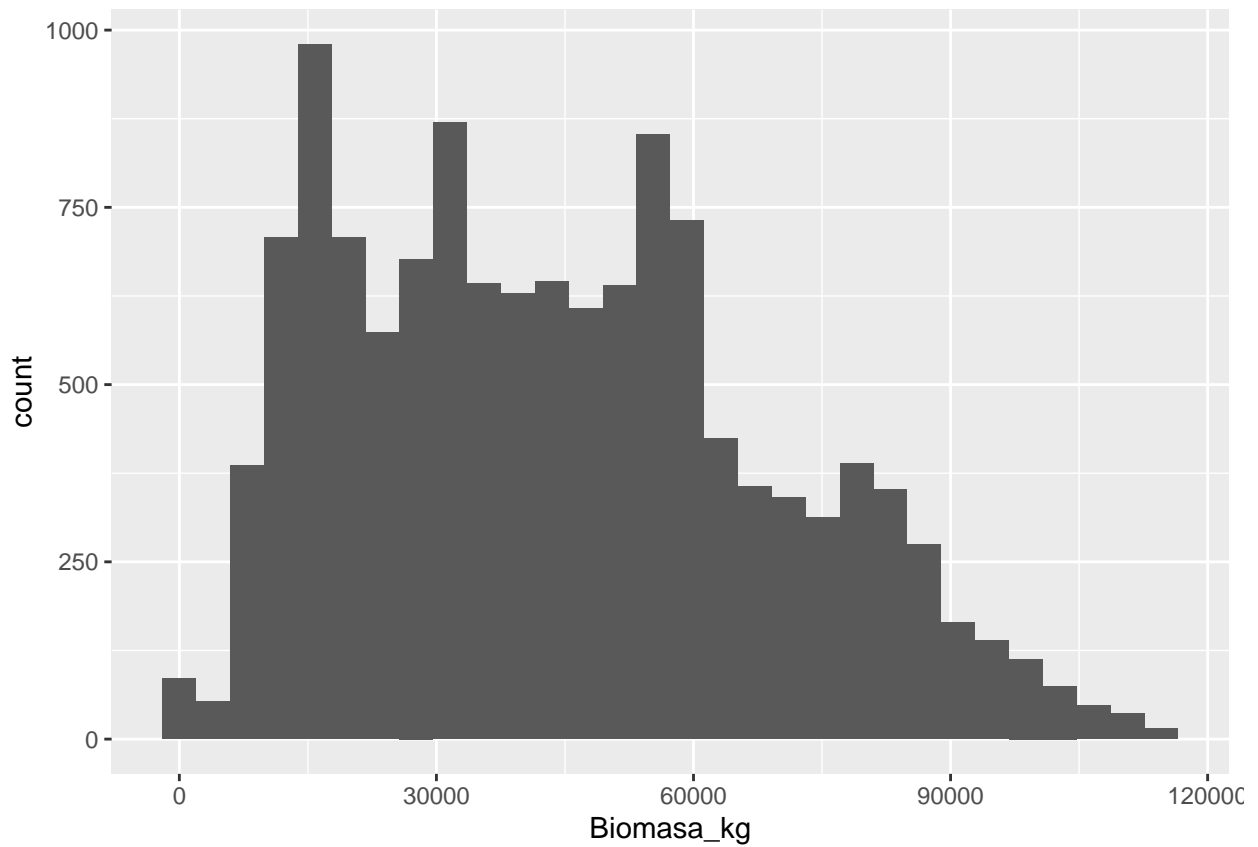
Densidad empírica

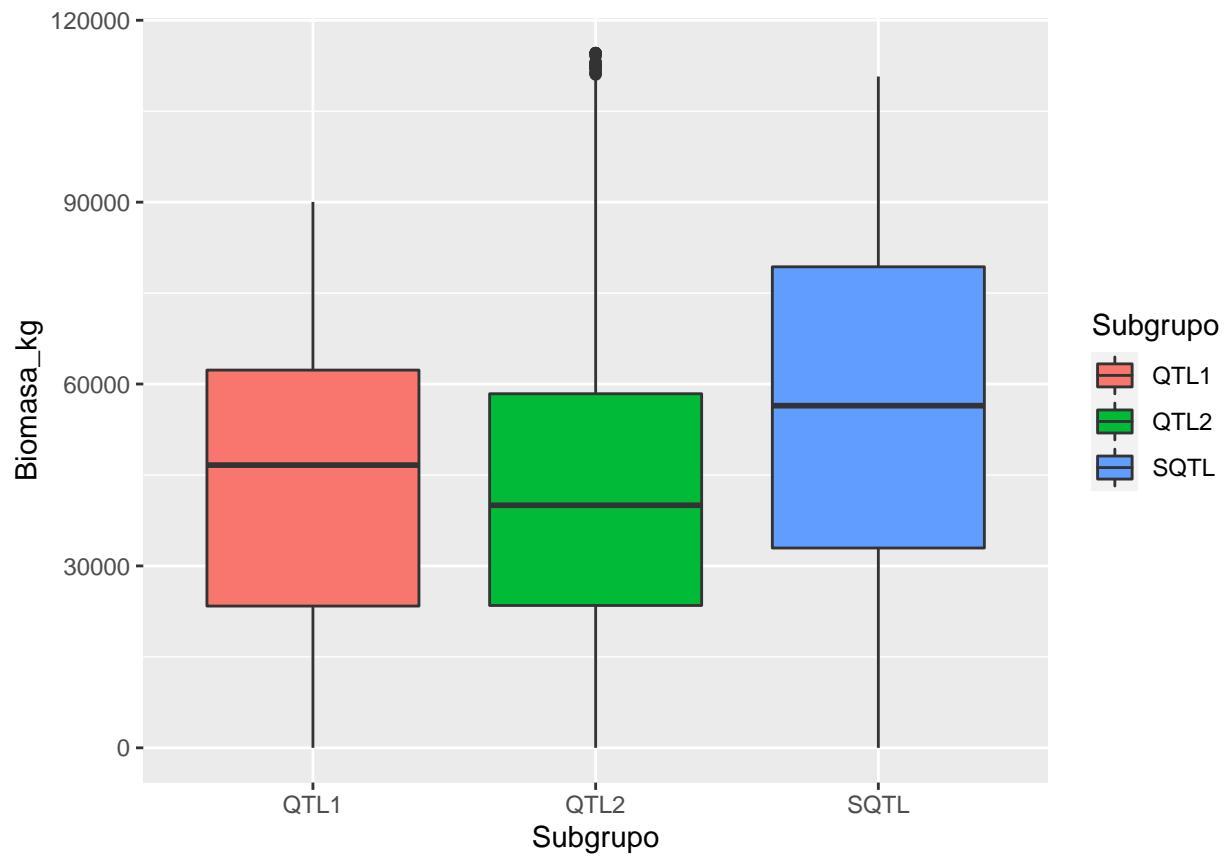


Distribución acumulada empírica

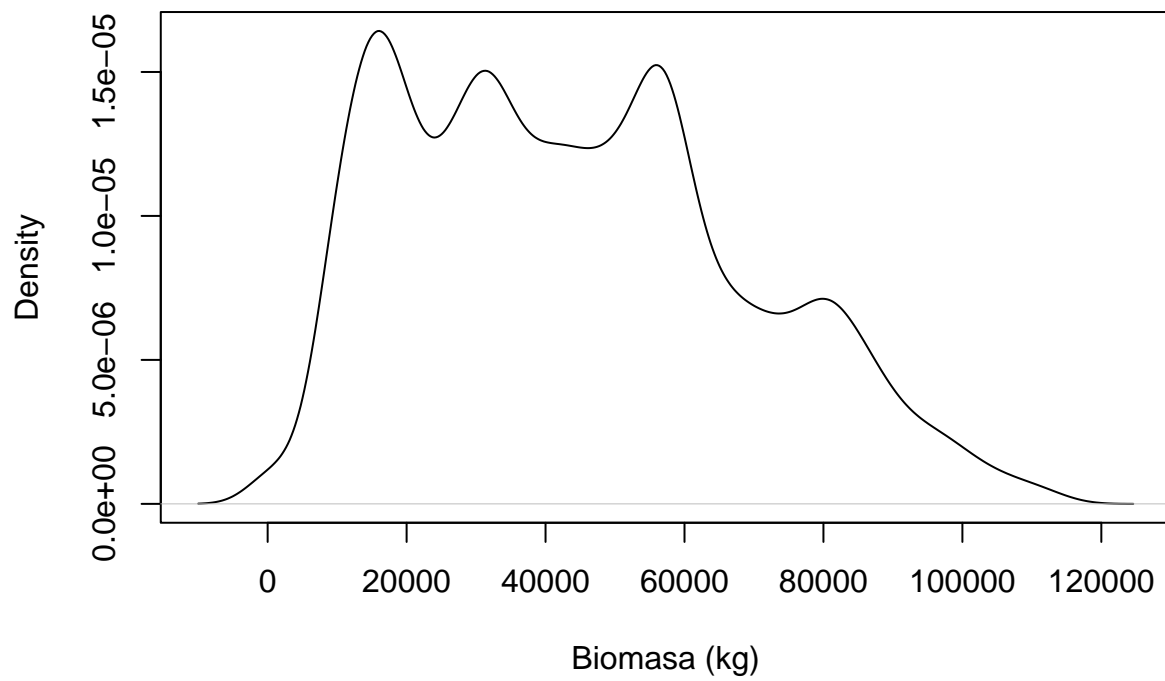


Biomasa viva por subgrupo

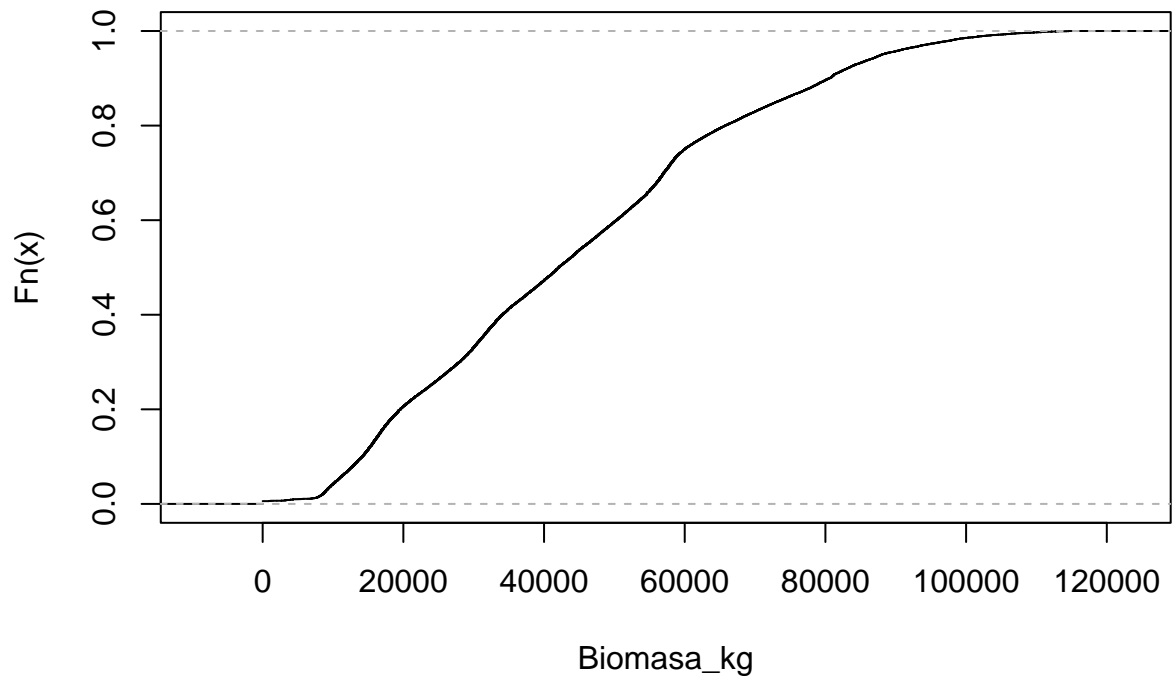




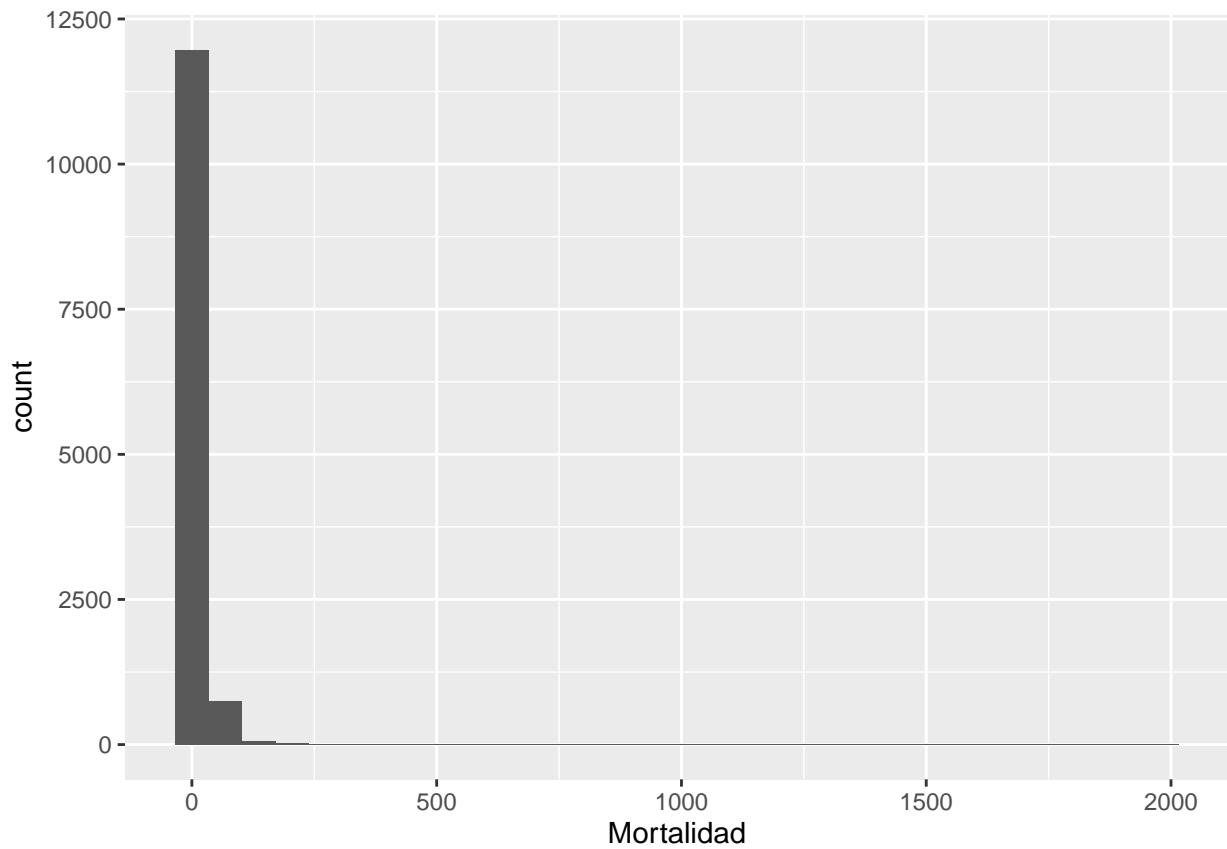
Densidad empírica



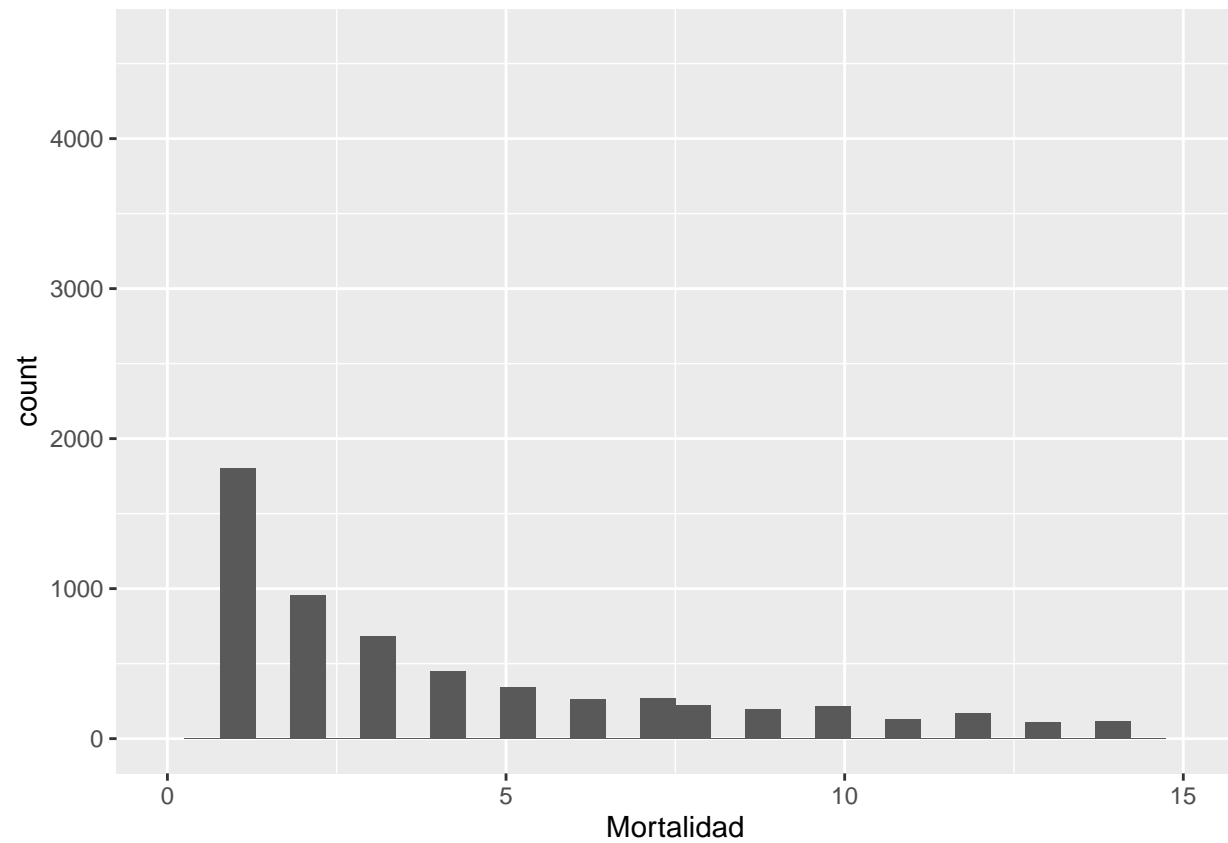
Distribución acumulada empírica

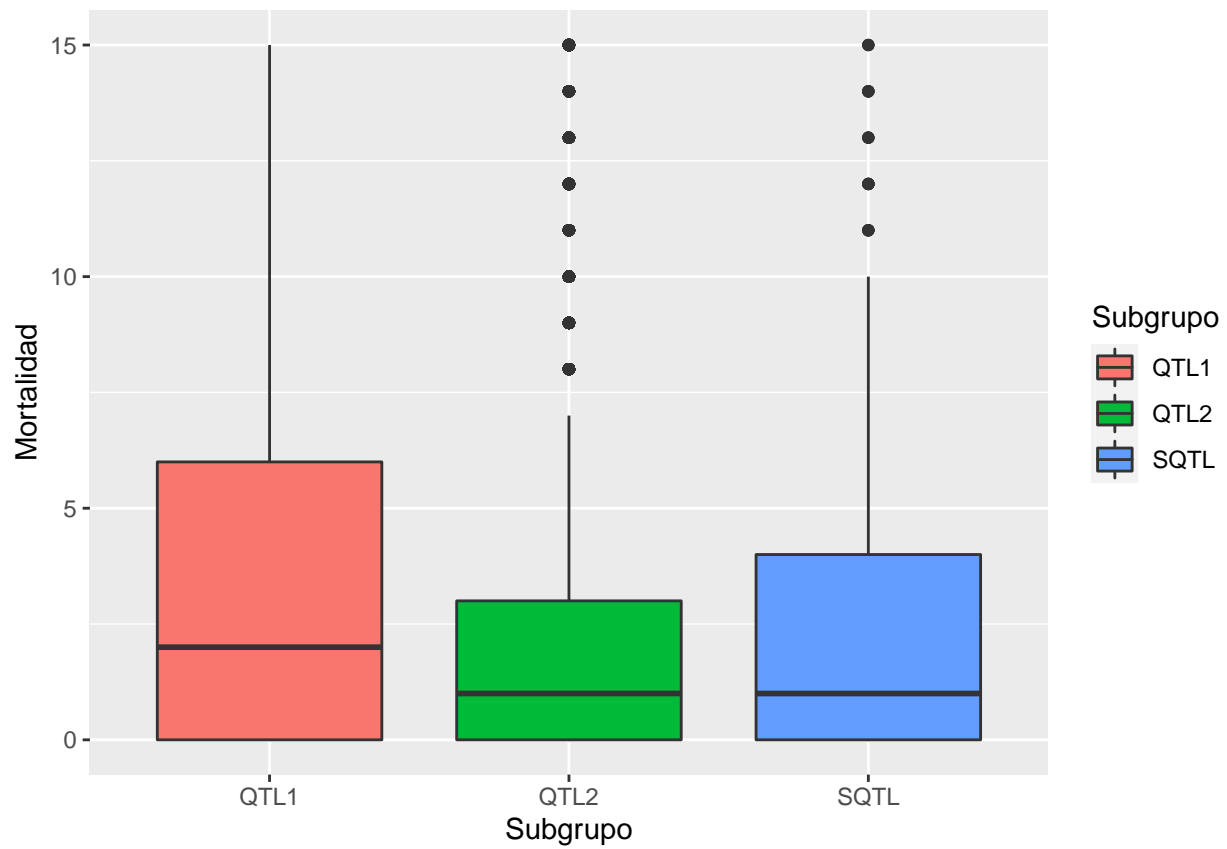


Mortalidad (n° peces) por subgrupo

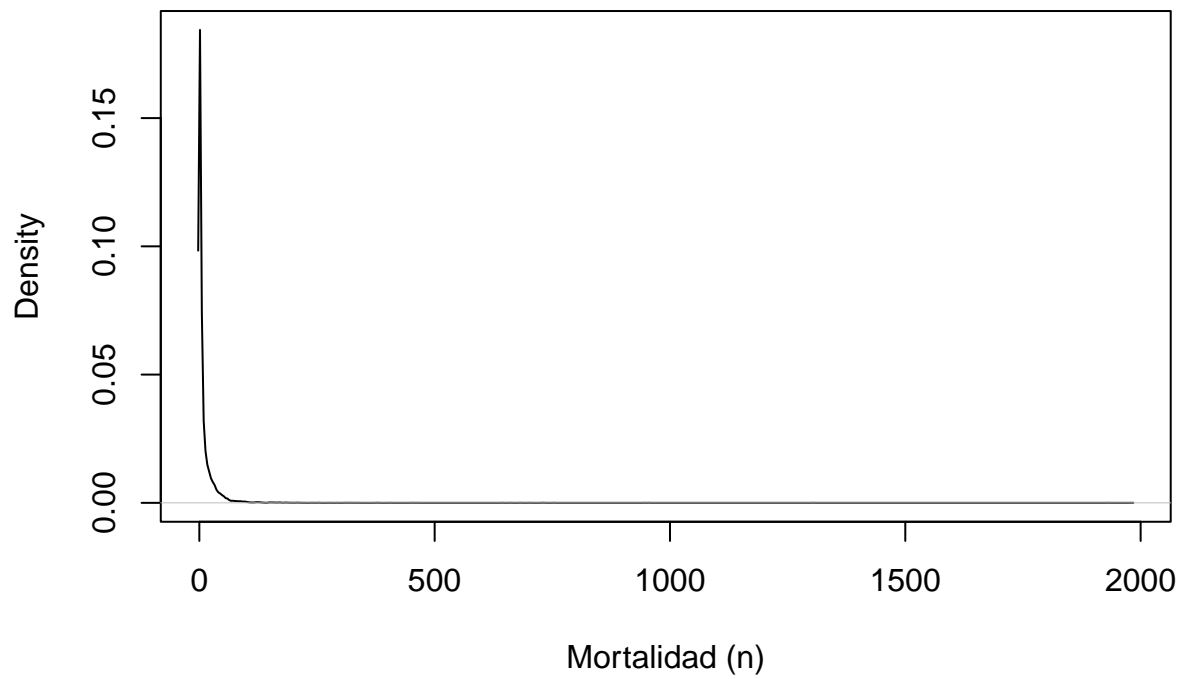


Para apreciar mejor como se comporta esta variable, haremos un histograma ajustando el eje x

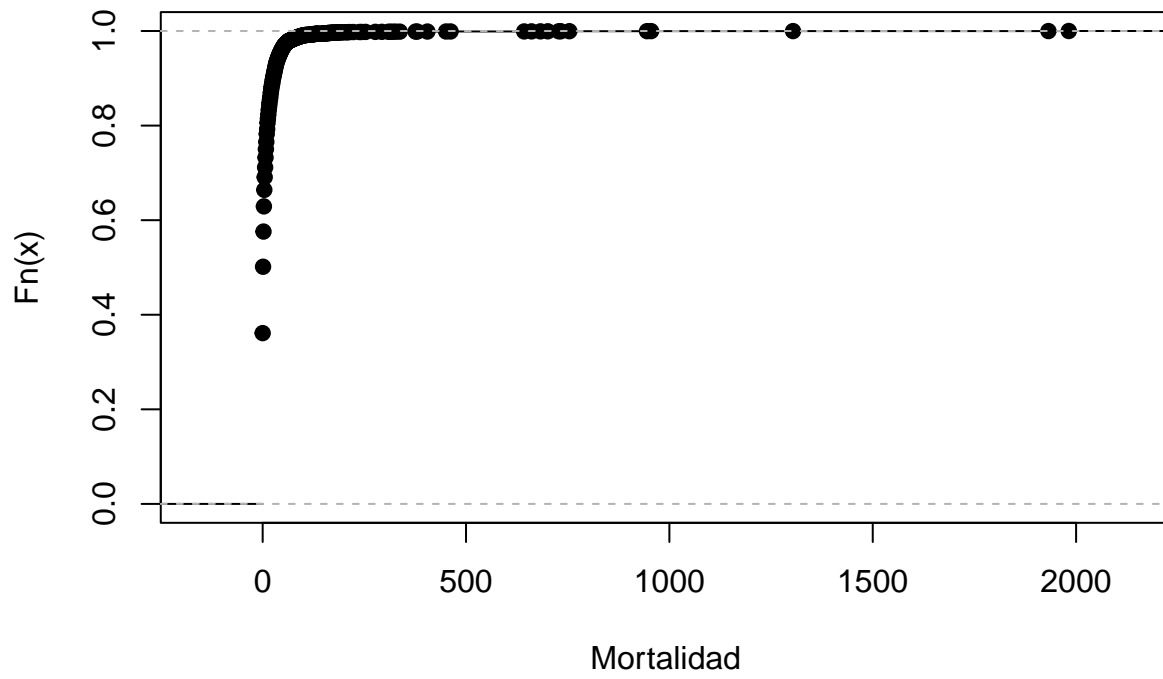




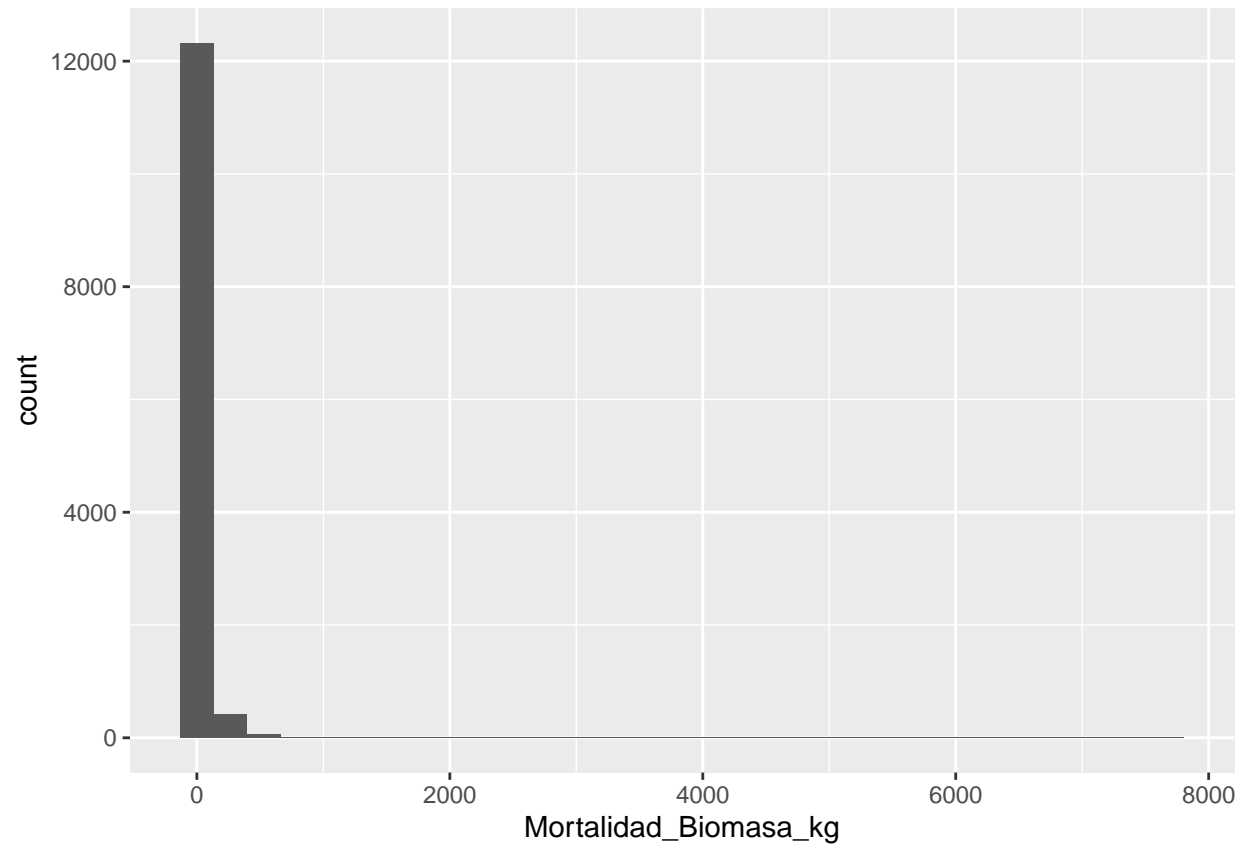
Densidad empírica

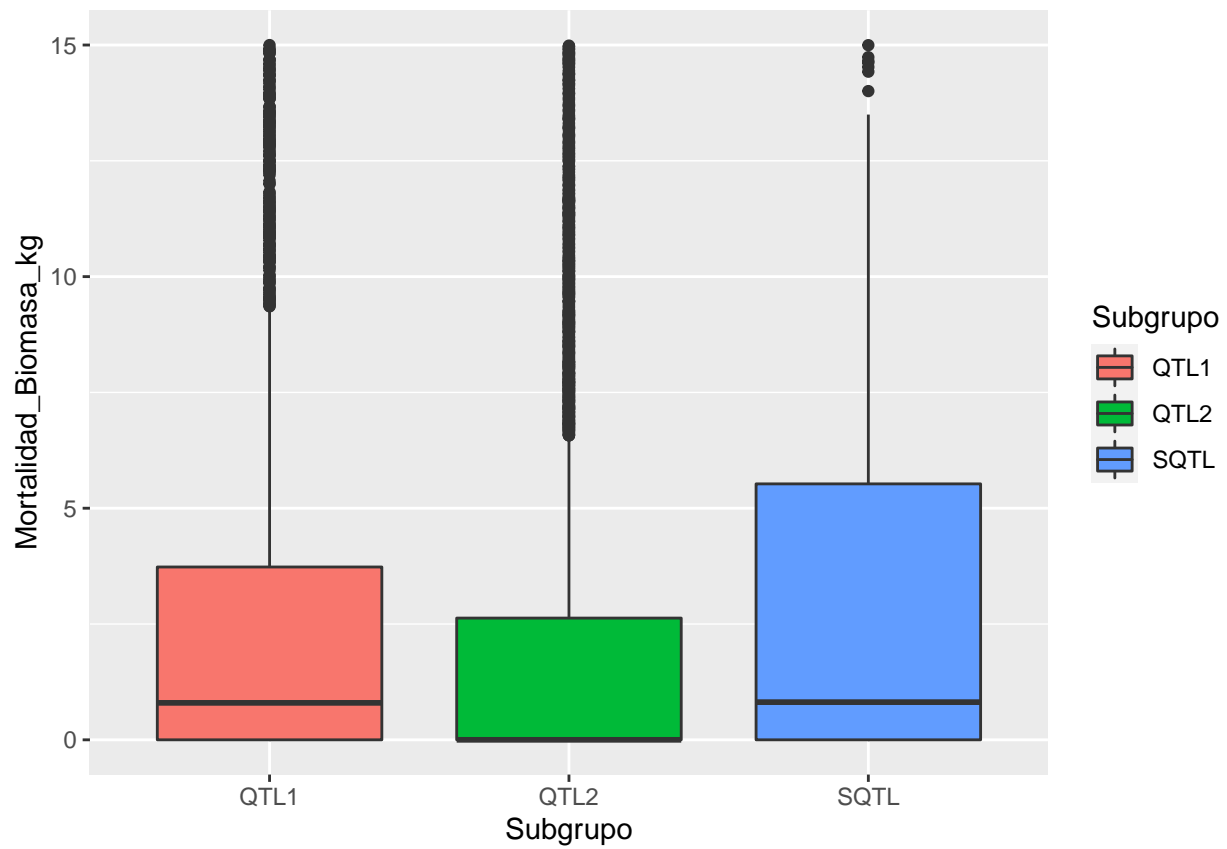


Distribución acumulada empírica

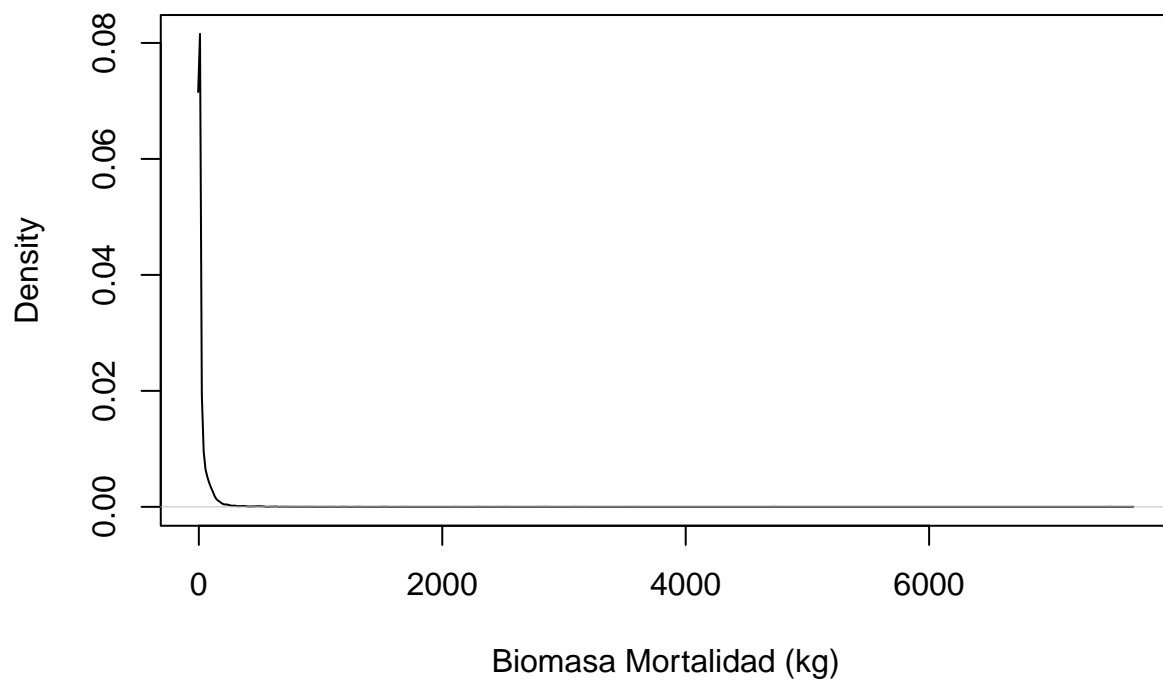


Biomasa de la mortalidad por subgrupo

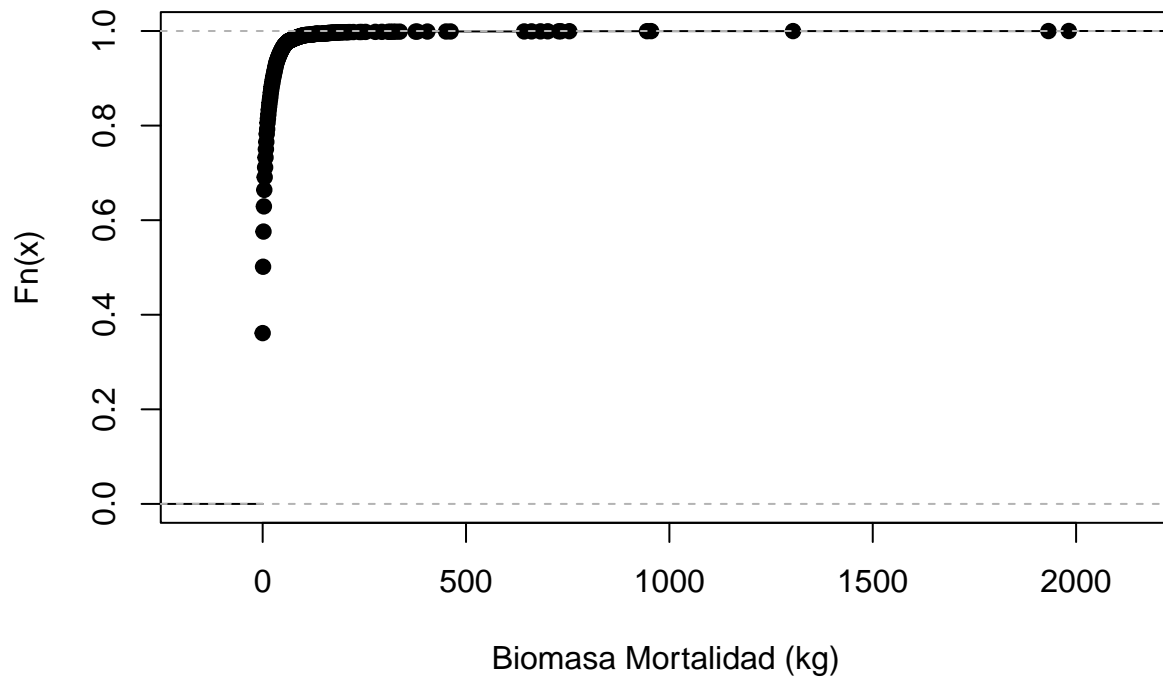




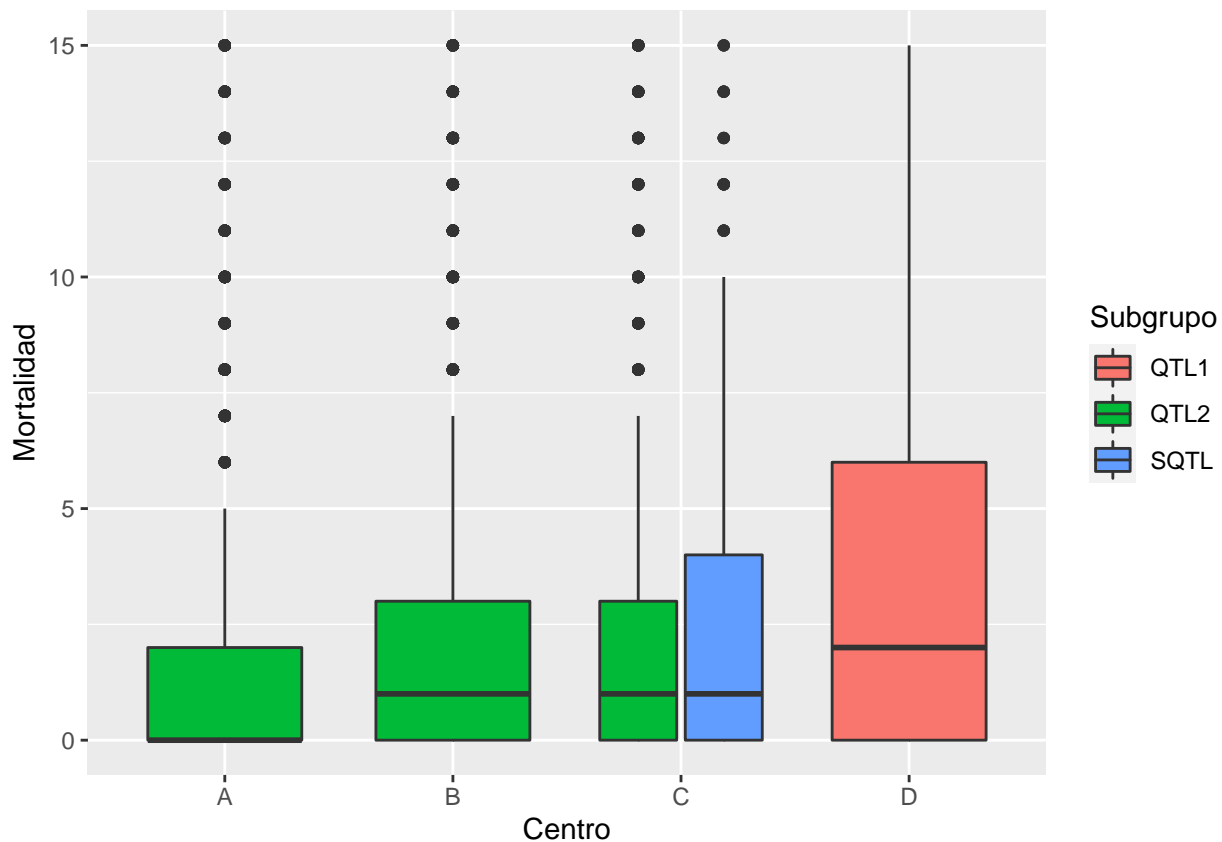
Densidad empírica



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.

Según lo sugerido en la evaluación 1, se agregan valores de estadística descriptiva al set de datos en formato de tabla.

Table 2: Media y varianza Mortalidad

Media_M	Var_M	sd	n
9.889123	1590.286	39.87839	12825

Table 3: Media y varianza Mortalidad por Subgrupo

Subgrupo	Media_M	Var_M	sd	n
QTL1	11.124386	880.1255	29.66691	3256
QTL2	9.599251	1923.0556	43.85266	9078
SQTL	7.057027	130.1478	11.40823	491

Table 4: Media y varianza Mortalidad por Centro y Subgrupo

Centro	Subgrupo	Media_M	Var_M	sd	n
A	QTL2	9.394431	3231.8437	56.84931	4130
B	QTL2	8.299193	403.1877	20.07953	3222
C	QTL2	12.516222	1618.7368	40.23353	1726
C	SQTL	7.057027	130.1478	11.40823	491
D	QTL1	11.124386	880.1255	29.66691	3256

Table 5: Media y varianza Mortalidad por Centro y Subgrupo

Centro	Subgrupo	Unidad	Media_M	Var_M	sd	n
A	QTL2	101	13.403061	20147.739273	141.942732	196
A	QTL2	102	7.873239	551.026309	23.473950	213
A	QTL2	103	4.208122	323.104423	17.975106	197
A	QTL2	104	11.328639	2765.985827	52.592640	213
A	QTL2	105	11.015544	8856.536216	94.109172	193
A	QTL2	106	10.230047	984.951546	31.383938	213
A	QTL2	107	4.693467	161.627785	12.713292	199
A	QTL2	108	5.751174	143.668926	11.986197	213
A	QTL2	109	14.210000	18670.227035	136.639039	200
A	QTL2	110	7.055814	192.426777	13.871798	215
A	QTL2	201	7.679245	504.541179	22.461994	212
A	QTL2	202	9.043269	2961.123723	54.416208	208
A	QTL2	203	10.638889	919.255039	30.319219	216
A	QTL2	204	9.467980	2340.675950	48.380533	203

Centro	Subgrupo	Unidad	Media_M	Var_M	sd	n
A	QTL2	205	13.154206	1223.342306	34.976311	214
A	QTL2	206	8.960784	858.560031	29.301195	204
A	QTL2	207	8.766667	464.533812	21.553047	210
A	QTL2	208	11.492462	1683.281509	41.027814	199
A	QTL2	209	8.309859	510.158251	22.586683	213
A	QTL2	210	10.854271	2668.115020	51.653800	199
B	QTL2	101	5.652361	99.486384	9.974286	233
B	QTL2	102	9.584746	705.545979	26.562116	236
B	QTL2	103	8.679654	255.870845	15.995963	231
B	QTL2	104	9.004202	446.569602	21.132193	238
B	QTL2	105	11.317391	822.339871	28.676469	230
B	QTL2	106	9.072340	489.340898	22.121051	235
B	QTL2	107	9.899123	578.390660	24.049754	228
B	QTL2	108	9.000000	386.364407	19.656154	237
B	QTL2	109	14.745000	1180.874347	34.363852	200
B	QTL2	110	6.793103	268.762203	16.393968	232
B	QTL2	111	5.352174	94.211676	9.706270	230
B	QTL2	112	6.628205	214.363321	14.641152	234
B	QTL2	113	5.784141	72.347004	8.505704	227
B	QTL2	114	5.489178	65.181404	8.073500	231
C	QTL2	101	10.946078	4617.795108	67.954361	204
C	QTL2	102	8.965174	1163.873781	34.115594	201
C	QTL2	103	12.748899	2555.613660	50.553078	227
C	QTL2	104	11.013699	788.976876	28.088732	219
C	QTL2	105	1.333333	5.333333	2.309401	3
C	QTL2	106	10.413793	1030.134858	32.095714	203
C	QTL2	108	15.748792	1480.383190	38.475748	207
C	QTL2	109	13.872428	371.186137	19.266192	243
C	QTL2	110	16.041096	1190.535001	34.504130	219
C	SQTL	105	3.479508	24.192994	4.918637	244
C	SQTL	107	10.465306	209.462931	14.472834	245
C	SQTL	109	26.000000	98.000000	9.899495	2
D	QTL1	101	10.483721	2563.297631	50.629020	215
D	QTL1	102	8.100840	149.761940	12.237726	238
D	QTL1	103	8.965368	186.720535	13.664572	231
D	QTL1	104	10.091304	270.467610	16.445899	230
D	QTL1	105	11.589958	317.982420	17.832062	239
D	QTL1	106	8.430962	266.002567	16.309585	239
D	QTL1	107	12.643478	692.815569	26.321390	230
D	QTL1	108	7.405530	277.714414	16.664766	217
D	QTL1	109	15.381148	355.191577	18.846527	244
D	QTL1	110	11.929515	666.384390	25.814422	227
D	QTL1	111	15.049587	3793.300435	61.589775	242
D	QTL1	112	11.571429	289.832429	17.024466	238
D	QTL1	113	10.489540	2009.015646	44.822044	239
D	QTL1	114	13.140969	451.564111	21.250038	227

FIN DE CURSO. ANALISIS ESTADISTICO

Se agrega una nueva variable al dataset (Days), la cual da cuenta en el tiempo del comportamiento de la variable respuesta (Mortalidad).

```
##      Fecha                Days                Centro
## Min.   :2019-02-09 00:00:00 Min.   : 1.0 Length:12821
## 1st Qu.:2019-06-01 00:00:00 1st Qu.: 61.0 Class :character
## Median :2019-07-26 00:00:00 Median :119.0 Mode  :character
## Mean   :2019-07-25 17:42:16 Mean   :116.5
## 3rd Qu.:2019-09-20 00:00:00 3rd Qu.:171.0
## Max.   :2019-12-09 00:00:00 Max.   :247.0
##
##      Unidad                Subgrupo                N_peces                Biomasa_kg
## Length:12821 Length:12821 Min.   : 0 Min.   : 0
## Class :character Class :character 1st Qu.:16308 1st Qu.: 23827
## Mode  :character Mode  :character Median :28345 Median : 42060
## Mean   :24150 Mean   : 44374
## 3rd Qu.:29029 3rd Qu.: 59946
## Max.   :29209 Max.   :114599
##
##      Mortalidad Mortalidad_Biomasa_kg N_cosecha Biomasa_cosecha
## Min.   : 0.000 Min.   : 0.000 Min.   : 0.00 Min.   : 0.0
## 1st Qu.: 0.000 1st Qu.: 0.000 1st Qu.: 0.00 1st Qu.: 0.0
## Median : 1.000 Median : 2.017 Median : 0.00 Median : 0.0
## Mean   : 9.892 Mean   : 27.558 Mean   : 76.92 Mean   : 270.3
## 3rd Qu.: 8.000 3rd Qu.: 17.845 3rd Qu.: 0.00 3rd Qu.: 0.0
## Max.   :1982.000 Max.   :7671.965 Max.   :22602.00 Max.   :78009.3
## NA's :4
##      Alimento_kg Temperatura
## Min.   : 0.0 Min.   : 0.00
## 1st Qu.: 285.0 1st Qu.:10.79
## Median : 446.0 Median :11.20
## Mean   : 440.8 Mean   :11.16
## 3rd Qu.: 604.0 3rd Qu.:11.70
## Max.   :1560.0 Max.   :15.00
##
## tibble [12,821 x 13] (S3: tbl_df/tbl/data.frame)
## $ Fecha      : POSIXct[1:12821], format: "2019-08-24" "2019-08-25" ...
## $ Days       : num [1:12821] 1 2 3 4 5 6 7 8 9 10 ...
## $ Centro     : chr [1:12821] "A" "A" "A" "A" ...
## $ Unidad     : chr [1:12821] "205" "205" "205" "205" ...
## $ Subgrupo   : chr [1:12821] "QTL2" "QTL2" "QTL2" "QTL2" ...
## $ N_peces    : num [1:12821] 16302 16302 16302 16302 16302 ...
## $ Biomasa_kg : num [1:12821] 35625 35879 36261 36684 37100 ...
## $ Mortalidad : num [1:12821] 0 0 0 0 0 0 0 1 2 2 ...
## $ Mortalidad_Biomasa_kg: num [1:12821] 0 0 0 0 0 ...
## $ N_cosecha  : num [1:12821] 0 0 0 0 0 0 0 0 0 0 ...
## $ Biomasa_cosecha : num [1:12821] 0 0 0 0 0 0 0 0 0 0 ...
## $ Alimento_kg : num [1:12821] 455 300 450 500 490 396 472 562 435 556 ...
## $ Temperatura : num [1:12821] 11.3 11.1 11.1 11.3 10.3 ...
```


Transformamos a factor lo que haya que transformar.

```
## tibble [12,821 x 13] (S3: tbl_df/tbl/data.frame)
## $ Fecha          : POSIXct[1:12821], format: "2019-08-24" "2019-08-25" ...
## $ Days           : num [1:12821] 1 2 3 4 5 6 7 8 9 10 ...
## $ 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 [1:12821] 16302 16302 16302 16302 16302 ...
## $ Biomasa_kg     : num [1:12821] 35625 35879 36261 36684 37100 ...
## $ Mortalidad     : num [1:12821] 0 0 0 0 0 0 0 1 2 2 ...
## $ Mortalidad_Biomasa_kg: num [1:12821] 0 0 0 0 0 ...
## $ N_cosecha      : num [1:12821] 0 0 0 0 0 0 0 0 0 0 ...
## $ Biomasa_cosecha : num [1:12821] 0 0 0 0 0 0 0 0 0 0 ...
## $ Alimento_kg    : num [1:12821] 455 300 450 500 490 396 472 562 435 556 ...
## $ Temperatura    : num [1:12821] 11.3 11.1 11.1 11.3 10.3 ...

##      Fecha          Days          Centro          Unidad
## Min.   :2019-02-09 00:00:00 Min.   : 1.0   A:4128   105   : 909
## 1st Qu.:2019-06-01 00:00:00 1st Qu.: 61.0   B:3221   107   : 902
## Median :2019-07-26 00:00:00 Median :119.0   C:2217   104   : 900
## Mean   :2019-07-25 17:42:16 Mean   :116.5   D:3255   110   : 892
## 3rd Qu.:2019-09-20 00:00:00 3rd Qu.:171.0           106   : 889
## Max.   :2019-12-09 00:00:00 Max.   :247.0           109   : 889
##                                     (Other):7440
## Subgrupo      N_peces      Biomasa_kg      Mortalidad
## QTL1:3255 Min.   : 0 Min.   : 0 Min.   : 0.000
## QTL2:9075 1st Qu.:16308 1st Qu.: 23827 1st Qu.: 0.000
## SQTL: 491 Median :28345 Median : 42060 Median : 1.000
##           Mean   :24150 Mean   : 44374 Mean   : 9.892
##           3rd Qu.:29029 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.   : 0.00 Min.   : 0.0 Min.   : 0.0
## 1st Qu.: 0.000 1st Qu.: 0.00 1st Qu.: 0.0 1st Qu.: 285.0
## Median : 2.017 Median : 0.00 Median : 0.0 Median : 446.0
## Mean   : 27.558 Mean   : 76.92 Mean   : 270.3 Mean   : 440.8
## 3rd Qu.: 17.845 3rd Qu.: 0.00 3rd Qu.: 0.0 3rd Qu.: 604.0
## Max.   :7671.965 Max.   :22602.00 Max.   :78009.3 Max.   :1560.0
##                                     NA's      :4
## Temperatura
## Min.   : 0.00
## 1st Qu.:10.79
## Median :11.20
## Mean   :11.16
## 3rd Qu.:11.70
## Max.   :15.00
##
```

Exploramos el comportamiento de la variable respuesta en función del tiempo (Days)

```
## Warning: Use of `dat2$Days` is discouraged. Use `Days` instead.
```

```
## Warning: Use of `dat2$Mortalidad` is discouraged. Use `Mortalidad` instead.  
## Warning: Use of `dat2$Days` is discouraged. Use `Days` instead.  
## Warning: Use of `dat2$Mortalidad` is discouraged. Use `Mortalidad` instead.  
## `geom_smooth()` using formula 'y ~ x'  
## Warning: Removed 12450 rows containing non-finite values (stat_smooth).  
## Warning: Removed 12450 rows containing missing values (geom_point).
```

Mortalidad vs Tiempo

