

# TRABAJO DIPLOMADO - Análisis Exploratorio

Marco Antonio Imués Figueroa

09 de mayo de 2022

## Presentación de tablas de datos

```
Chlorella <- read_excel("Chlorella.xlsx", na="NA")
Chlorella_mod <- na.omit(Chlorella)
Chlorella_mod$FBR <- as.factor(Chlorella_mod$FBR)
Chlorella_mod$Ciclo <- as.factor(Chlorella_mod$Ciclo)
Chlorella_mod$Dias <- as.factor(Chlorella_mod$Dias)
Chlorella_mod$Muestra <- as.factor(Chlorella_mod$Muestra)
summary(Chlorella_mod)
```

```
##           FBR      Ciclo      Dias      Muestra      Densidad
## Conico      :90    1:63    2      :36    1:80    Min.   : 102000
## Helicoidal:81    2:87    3      :36    2:80    1st Qu.: 632708
## Serpentin :69    3:45    4      :36    3:80    Median : 2291666
##           4:45    5      :36           Mean   : 4474614
##           6      :36           3rd Qu.: 5454167
##           7      :18           Max.    :24600000
##           (Other):42
##           TCS      Temperatura      CO2      OD
## Min.    :-0.09781  Min.    :19.00  Min.    :0.1600  Min.    :3.120
## 1st Qu.: 0.00555  1st Qu.:23.40  1st Qu.:0.3800  1st Qu.:4.058
## Median : 0.02008  Median :24.00  Median :0.4600  Median :4.400
## Mean    : 0.01822  Mean    :24.05  Mean    :0.5188  Mean    :4.321
## 3rd Qu.: 0.03232  3rd Qu.:24.60  3rd Qu.:0.6000  3rd Qu.:4.640
## Max.    : 0.08177  Max.    :27.80  Max.    :1.7600  Max.    :6.910
##
```

```
str(Chlorella_mod)
```

```
## tibble [240 × 9] (S3: tbl_df/tbl/data.frame)
## $ FBR      : Factor w/ 3 levels "Conico","Helicoidal",...: 2 2 2 2 2 2 2 2 2 2 ...
## $ Ciclo    : Factor w/ 4 levels "1","2","3","4": 1 1 1 1 1 1 1 1 1 1 ...
## $ Dias     : Factor w/ 13 levels "2","3","4","5",...: 1 1 1 2 2 2 3 3 3 4 ...
## $ Muestra  : Factor w/ 3 levels "1","2","3": 1 2 3 1 2 3 1 2 3 1 ...
## $ Densidad : num [1:240] 476667 483333 500000 663333 943333 ...
## $ TCS      : num [1:240] 0.0496 0.0502 0.0516 0.0138 0.0279 ...
## $ Temperatura: num [1:240] 24.1 24.6 24.2 23.8 25.3 25.9 24.2 27.8 26.5 22.8 ...
## $ CO2      : num [1:240] 0.4 0.4 0.4 0.28 0.28 0.28 0.3 0.3 0.3 0.4 ...
## $ OD       : num [1:240] 4.05 3.97 4 4.01 3.9 3.62 4.05 3.78 3.67 4.08 ...
## - attr(*, "na.action")= 'omit' Named int [1:36] 1 2 3 25 26 27 58 59 60 76 ...
## ... attr(*, "names")= chr [1:36] "1" "2" "3" "25" ...
```

```
table(Chlorella_mod$FBR)
```

```
##
##      Conico Helicoidal  Serpentin
##      90      81      69
```

```
table(Chlorella_mod$Ciclo)
```

```
##
##  1  2  3  4
## 63 87 45 45
```

```
table(Chlorella_mod$Dias)
```

```
##
##  2  3  4  5  6  7  8  9 10 11 12 13 14
## 36 36 36 36 36 18 15  6  6  6  3  3  3
```

```
table(Chlorella_mod$Muestra)
```

```
##
## 1 2 3
## 80 80 80
```

```
Chlorella_mod <- data.frame(Chlorella_mod)
knitr::kable(head(Chlorella_mod), digits = 6, caption = "Tabla 1. Resumen de variables en el cultivo de la microalga *Chlorella vulgaris* en FBRT")
```

Tabla 1. Resumen de variables en el cultivo de la microalga *Chlorella vulgaris* en FBRT

FBR	Ciclo	Dias	Muestra	Densidad	TCS	Temperatura	CO2	OD
Helicoidal	1	2	1	476667	0.049587	24.1	0.40	4.05
Helicoidal	1	2	2	483333	0.050166	24.6	0.40	3.97
Helicoidal	1	2	3	500000	0.051578	24.2	0.40	4.00
Helicoidal	1	3	1	663333	0.013769	23.8	0.28	4.01
Helicoidal	1	3	2	943333	0.027863	25.3	0.28	3.90
Helicoidal	1	3	3	605333	0.007965	25.9	0.28	3.62

```

Tabla_FBR <- Chlorella_mod %>% group_by(FBR) %>%
  summarize("Prom. Densidad"= mean(Densidad),
            "SD Densidad"=sd(Densidad),
            "Prom. TCS" = mean(TCS),
            "SD TCS" = sd(TCS),
            "Prom. Temp." = mean(Temperatura),
            "SD Temp." = sd(Temperatura),
            "Prom. OD" = mean (OD),
            "SD OD" = sd(OD),
            "Prom. CO2" = mean(CO2),
            "SD CO2" = sd(CO2),
            n=n ())

knitr::kable(Tabla_FBR, digits = 4, caption = "Tabla 2. Resumen de las variables en estudio por FBRT")

```

Tabla 2. Resumen de las variables en estudio por FBRT

<b>FBR</b>	<b>Prom. Densidad</b>	<b>SD Densidad</b>	<b>Prom. TCS</b>	<b>SD TCS</b>	<b>Prom. Temp.</b>	<b>SD Temp.</b>	<b>Prom. OD</b>	<b>SD OD</b>	<b>Prom. CO2</b>	<b>SD CO2</b>	<b>n</b>
Conico	6086504	7018610	0.0187	0.0201	23.5222	1.0798	4.3997	0.4999	0.4860	0.1521	90
Helicoidal	4907137	4886834	0.0223	0.0242	24.1185	1.2034	4.4599	0.4011	0.5563	0.2939	81
Serpentin	1864404	1573941	0.0128	0.0292	24.6609	0.5470	4.0557	0.4143	0.5174	0.2550	69

```

Tabla_ciclo <- Chlorella_mod %>% group_by(Ciclo) %>%
  summarize("Prom. Densidad"= mean(Densidad),
            "SD Densidad"=sd(Densidad),
            "Prom. TCS" = mean(TCS),
            "SD TCS" = sd(TCS),
            "Prom. Temp." = mean(Temperatura),
            "SD Temp." = sd(Temperatura),
            "Prom. OD" = mean (OD),
            "SD OD" = sd(OD),
            "Prom. CO2" = mean(CO2),
            "SD CO2" = sd(CO2),
            n=n ())

knitr::kable(Tabla_ciclo, digits = 4, caption = "Tabla 3. Resumen de las variables en estudio por ciclo de cultivo")

```

Tabla 3. Resumen de las variables en estudio por ciclo de cultivo

Ciclo	Prom. Densidad	SD Densidad	Prom. TCS	SD TCS	Prom. Temp.	SD Temp.	Prom. OD	SD OD	Prom. CO2	SD CO2	n
1	2760765	2134151	0.0157	0.0259	24.5778	1.1509	3.6973	0.2436	0.3895	0.1312	63
2	6036368	7417723	0.0135	0.0260	23.6678	1.1622	4.6038	0.3781	0.5221	0.1522	87
3	3922811	4528448	0.0249	0.0213	24.0444	0.5903	4.4631	0.1799	0.7133	0.3688	45
4	4406415	4406804	0.0241	0.0205	24.0600	1.0367	4.5058	0.2360	0.4987	0.2015	45

```

Tabla_FBR_ciclo <- Chlorella_mod %>% group_by(FBR, Ciclo) %>%
  summarize("Prom. Densidad"= mean(Densidad),
            "SD Densidad"=sd(Densidad),
            "Prom. TCS" = mean(TCS),
            "SD TCS" = sd(TCS),
            "Prom. Temp." = mean(Temperatura),
            "SD Temp." = sd(Temperatura),
            "Prom. OD" = mean (OD),
            "SD OD" = sd(OD),
            "Prom. CO2" = mean(CO2),
            "SD CO2" = sd(CO2),
            n=n ())

```

```

## `summarise()` has grouped output by 'FBR'. You can override using the `.groups`
## argument.

```

```

knitr::kable(Tabla_FBR_ciclo, digits = 4, caption = "Tabla 4. Resumen de las variables en estudio por FBR y ciclo de cultivo")

```

Tabla 4. Resumen de las variables en estudio por FBR y ciclo de cultivo

<b>FBR</b>	<b>Ciclo</b>	<b>Prom. Densidad</b>	<b>SD Densidad</b>	<b>Prom. TCS</b>	<b>SD TCS</b>	<b>Prom. Temp.</b>	<b>SD Temp.</b>	<b>Prom. OD</b>	<b>SD OD</b>	<b>Prom. CO2</b>	<b>SD CO2</b>	<b>n</b>
Conico	1	2098793.5	1338862.1	0.0125	0.0135	24.1524	1.3754	3.6924	0.1655	0.4400	0.1658	21
Conico	2	8887658.1	8751627.5	0.0149	0.0198	23.2462	1.0495	4.6985	0.4138	0.4646	0.1153	39
Conico	3	4568500.0	5088305.2	0.0268	0.0222	23.4600	0.3269	4.4973	0.1744	0.5320	0.1389	15
Conico	4	5904303.7	5348065.6	0.0290	0.0218	23.4200	0.8890	4.5153	0.1756	0.5600	0.1999	15
Helicoidal	1	4021565.7	2629916.3	0.0214	0.0268	24.8524	1.1707	3.9100	0.1601	0.3657	0.0810	21
Helicoidal	2	5497999.8	5841868.4	0.0157	0.0273	23.6100	1.2472	4.7133	0.2556	0.5940	0.1430	30
Helicoidal	3	5016221.9	5491982.8	0.0311	0.0143	24.2467	0.6323	4.5487	0.1966	0.8000	0.5291	15

FBR	Ciclo	Prom. Densidad	SD Densidad	Prom. TCS	SD TCS	Prom. Temp.	SD Temp.	Prom. OD	SD OD	Prom. CO2	SD CO2	n
Helicoidal	4	4856125.9	4877356.4	0.0278	0.0186	23.9800	1.1384	4.6340	0.2763	0.5040	0.1948	15
Serpentin	1	2161936.5	1704261.7	0.0132	0.0335	24.7286	0.7424	3.4895	0.1952	0.3629	0.1237	21
Serpentin	2	755851.8	492487.7	0.0069	0.0349	24.6778	0.4735	4.2161	0.1677	0.5267	0.1892	18
Serpentin	3	2183711.1	1851815.0	0.0168	0.0246	24.4267	0.1486	4.3433	0.0903	0.8080	0.2745	15
Serpentin	4	2458814.7	1399616.7	0.0156	0.0194	24.7800	0.5388	4.3680	0.1726	0.4320	0.2022	15

```

Tabla_FBR_muestra <- Chlorella_mod %>% group_by(FBR, Muestra) %>%
  summarize("Prom. Densidad"= mean(Densidad),
            "SD Densidad"=sd(Densidad),
            "Prom. TCS" = mean(TCS),
            "SD TCS" = sd(TCS),
            "Prom. Temp." = mean(Temperatura),
            "SD Temp." = sd(Temperatura),
            "Prom. OD" = mean (OD),
            "SD OD" = sd(OD),
            "Prom. CO2" = mean(CO2),
            "SD CO2" = sd(CO2),
            n=n ())

```

```

## `summarise()` has grouped output by 'FBR'. You can override using the `.groups`
## argument.

```

```

knitr::kable(Tabla_FBR_muestra, digits = 4, caption = "Tabla 5. Resumen de las variables en estudio por FBR y muestra")

```

Tabla 5. Resumen de las variables en estudio por FBR y muestra

FBR	Muestra	Prom. Densidad	SD Densidad	Prom. TCS	SD TCS	Prom. Temp.	SD Temp.	Prom. OD	SD OD	Prom. CO2	SD CO2	n
Conico	1	5707928	6925772	0.0186	0.0191	22.8867	1.1392	4.4360	0.3919	0.4860	0.1539	30

FBR	Muestra	Prom. Densidad	SD Densidad	Prom. TCS	SD TCS	Prom. Temp.	SD Temp.	Prom. OD	SD OD	Prom. CO2	SD CO2	n
Conico	2	6145052	7126774	0.0192	0.0203	24.0033	1.0692	4.4133	0.6398	0.4860	0.1539	30
Conico	3	6406533	7222534	0.0183	0.0216	23.6767	0.6750	4.3497	0.4472	0.4860	0.1539	30
Helicoidal	1	4592749	4632478	0.0220	0.0235	23.3148	1.3598	4.4826	0.3487	0.5563	0.2977	27
Helicoidal	2	4887479	4793282	0.0222	0.0225	24.5000	0.9307	4.4752	0.4506	0.5563	0.2977	27
Helicoidal	3	5241183	5368995	0.0227	0.0272	24.5407	0.8568	4.4219	0.4099	0.5563	0.2977	27
Serpentin	1	1566309	1359794	0.0101	0.0402	24.4783	0.4889	4.0917	0.3945	0.5174	0.2588	23
Serpentin	2	1955717	1629190	0.0146	0.0198	24.7652	0.6027	4.0452	0.4483	0.5174	0.2588	23
Serpentin	3	2071186	1734202	0.0138	0.0247	24.7391	0.5194	4.0300	0.4143	0.5174	0.2588	23

## Exploración gráfica:

### Histogramas



```

Chlorella_mod <- data.frame(Chlorella_mod)
hist1 <- ggplot(Chlorella_mod, aes(x = Densidad)) +
  geom_histogram(bins = 10, colour = "blue", alpha = 0.3) +
  labs(y = "Frecuencia")

hist2 <- ggplot(Chlorella_mod, aes(x = TCS)) +
  geom_histogram(bins = 10, na.rm = TRUE, colour = "blue", alpha = 0.3) +
  labs(y = "Frecuencia")

hist3 <- ggplot(Chlorella_mod, aes(x = Temperatura)) +
  geom_histogram(bins = 10, colour = "blue", alpha = 0.3) +
  labs(y = "Frecuencia")

hist4 <- ggplot(Chlorella_mod, aes(x = CO2)) +
  geom_histogram(bins = 10, colour = "blue", alpha = 0.3) +
  labs(y = "Frecuencia")

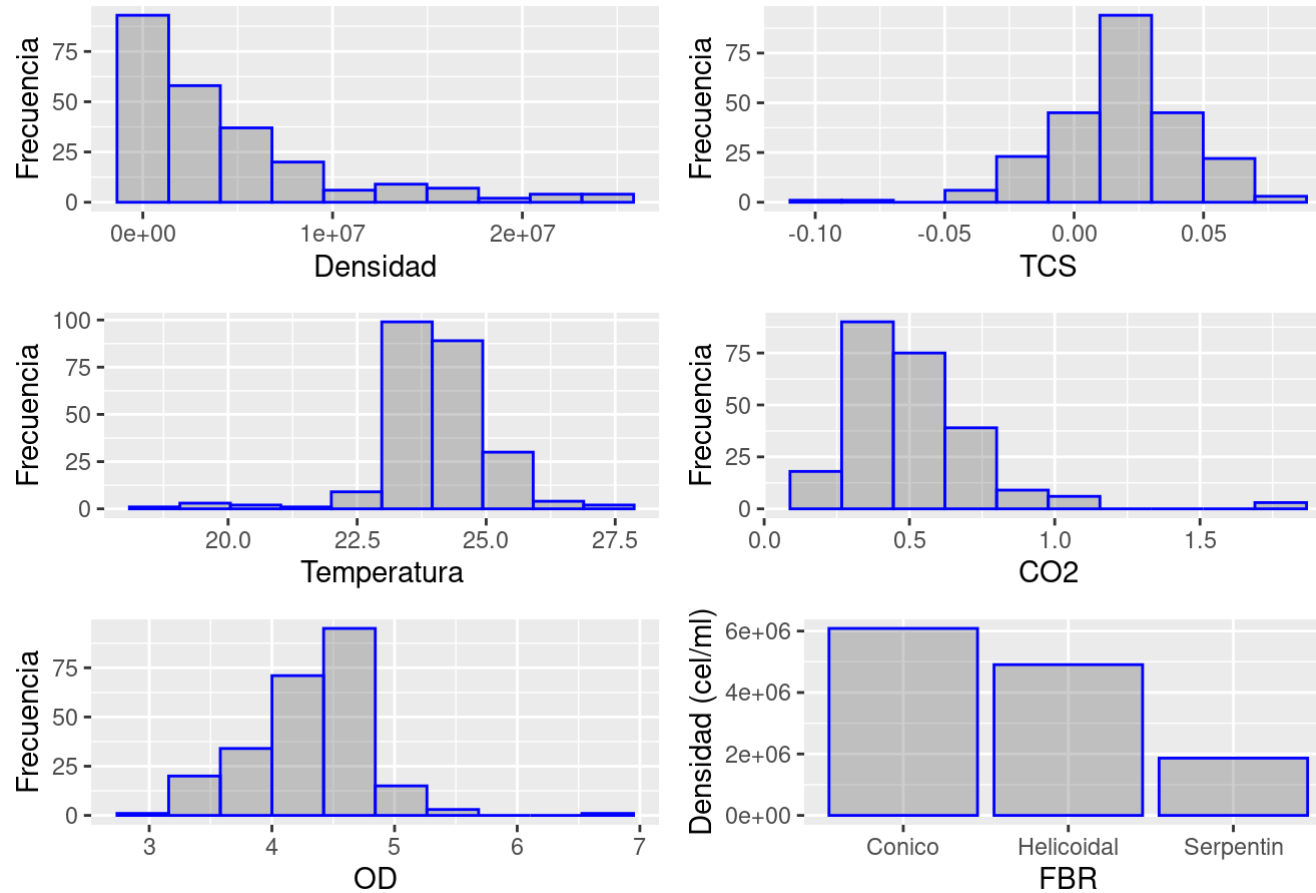
hist5 <- ggplot(Chlorella_mod, aes(x = OD)) +
  geom_histogram(bins = 10, colour="blue", alpha = 0.3) +
  labs(y = "Frecuencia")

hist6 <- ggplot(Tabla_FBR, aes(x = FBR, y = `Prom. Densidad`))+
  geom_col(colour="blue", alpha = 0.3) + labs(y = "Densidad (cel/ml)")

grid.arrange(hist1, hist2, hist3, hist4, hist5, hist6, ncol=2, nrow=3, top = textGrob("Figura 1. Histogramas para representa
r los valores de la diferentes variables", gp=gpar(fontsize=12, font=3)))

```

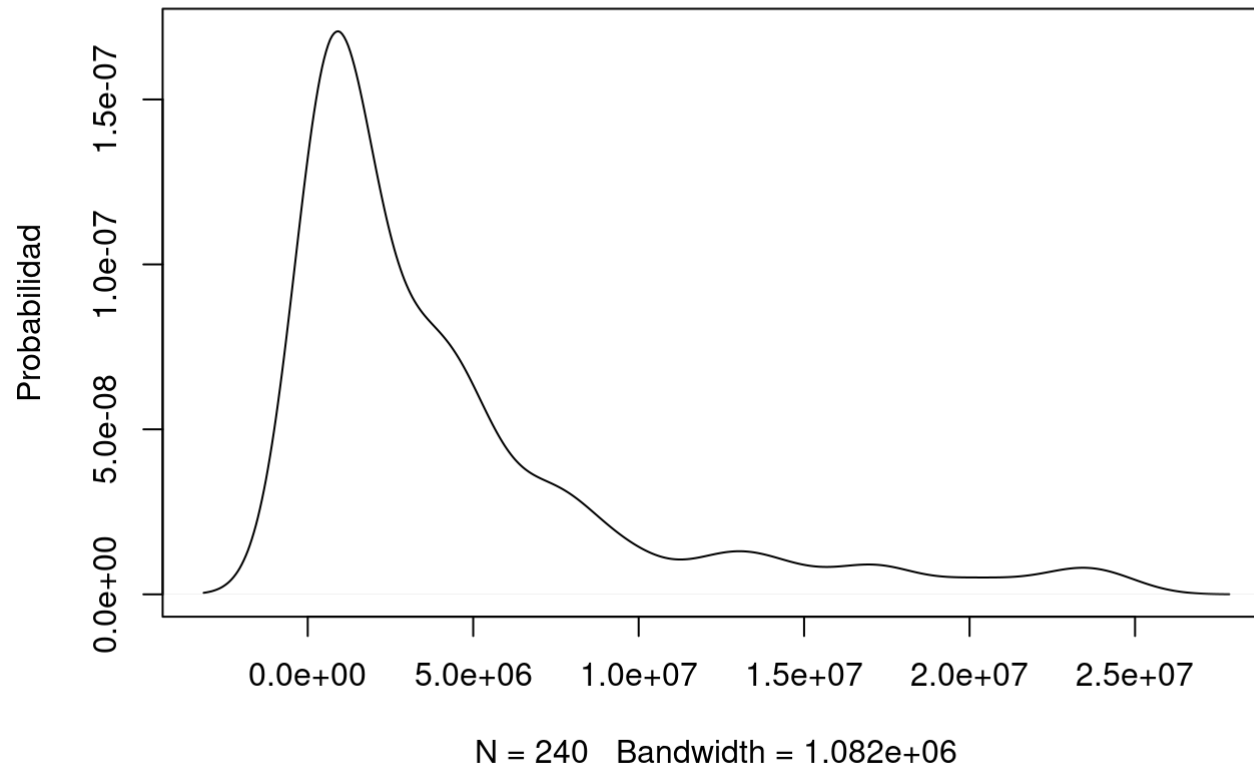
Figura 1. Histogramas para representar los valores de la diferentes variables



## Función de densidad y Función de probabilidad acumulada

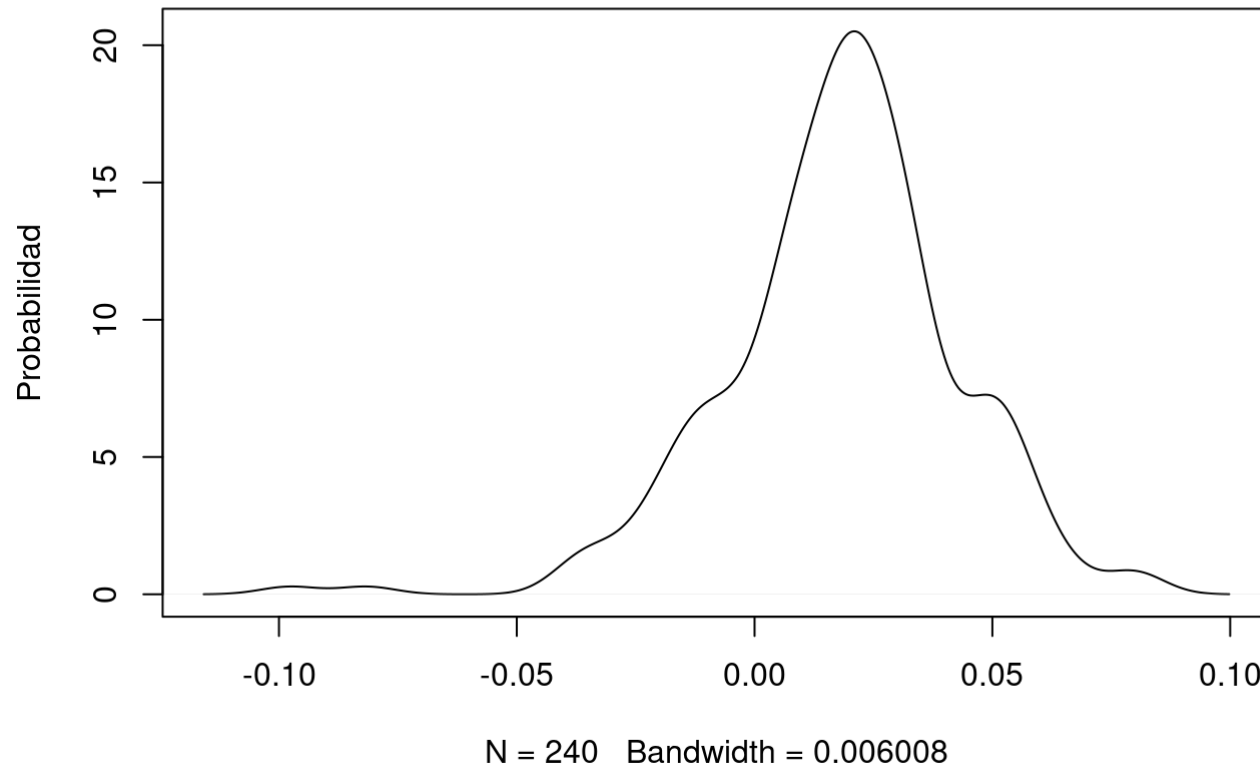
```
plot(density(Chlorella_mod$Densidad), main = "Figura 2. Función de densidad para la variable Densidad poblacional (cel/ml)",
y = "Probabilidad")
```

**Figura 2. Función de densidad para la variable Densidad poblacional (cel/**



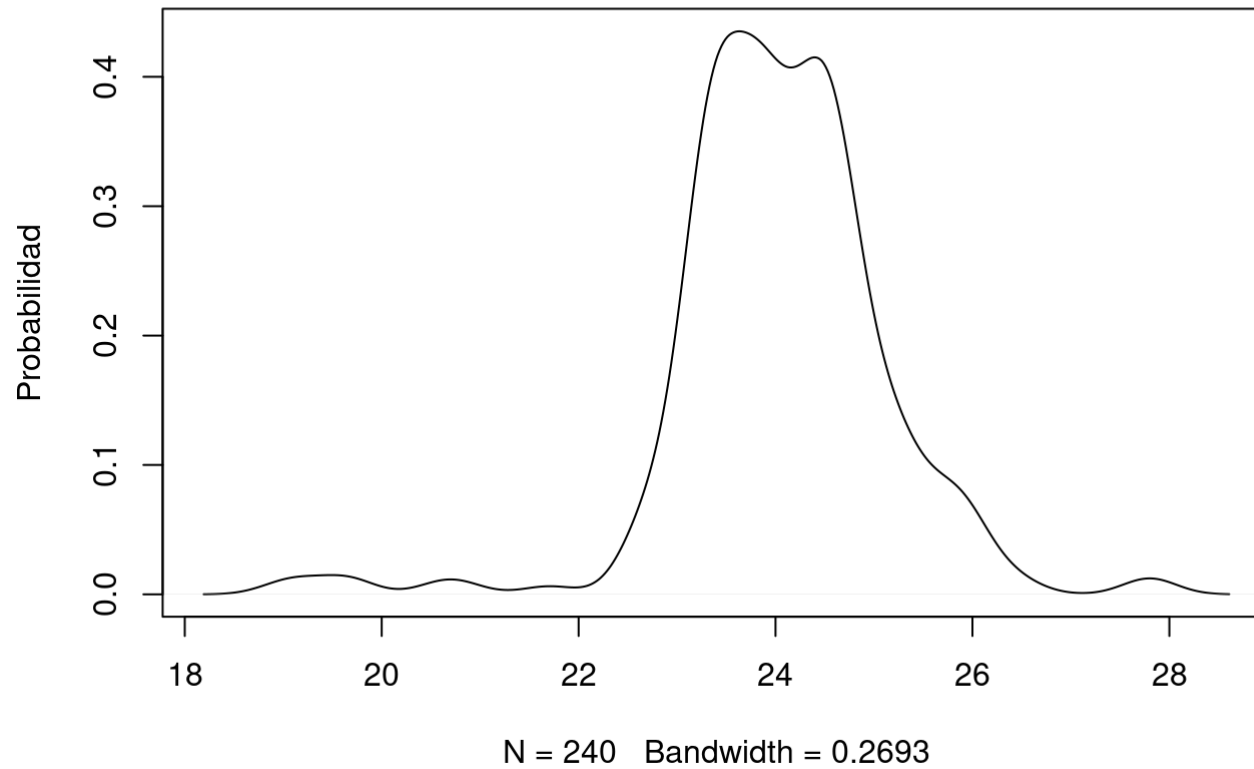
```
plot(density(Chlorella_mod$TCS), main = "Figura 3. Función de densidad para la variable TCS", y = "Probabilidad")
```

**Figura 3. Función de densidad para la variable TCS**



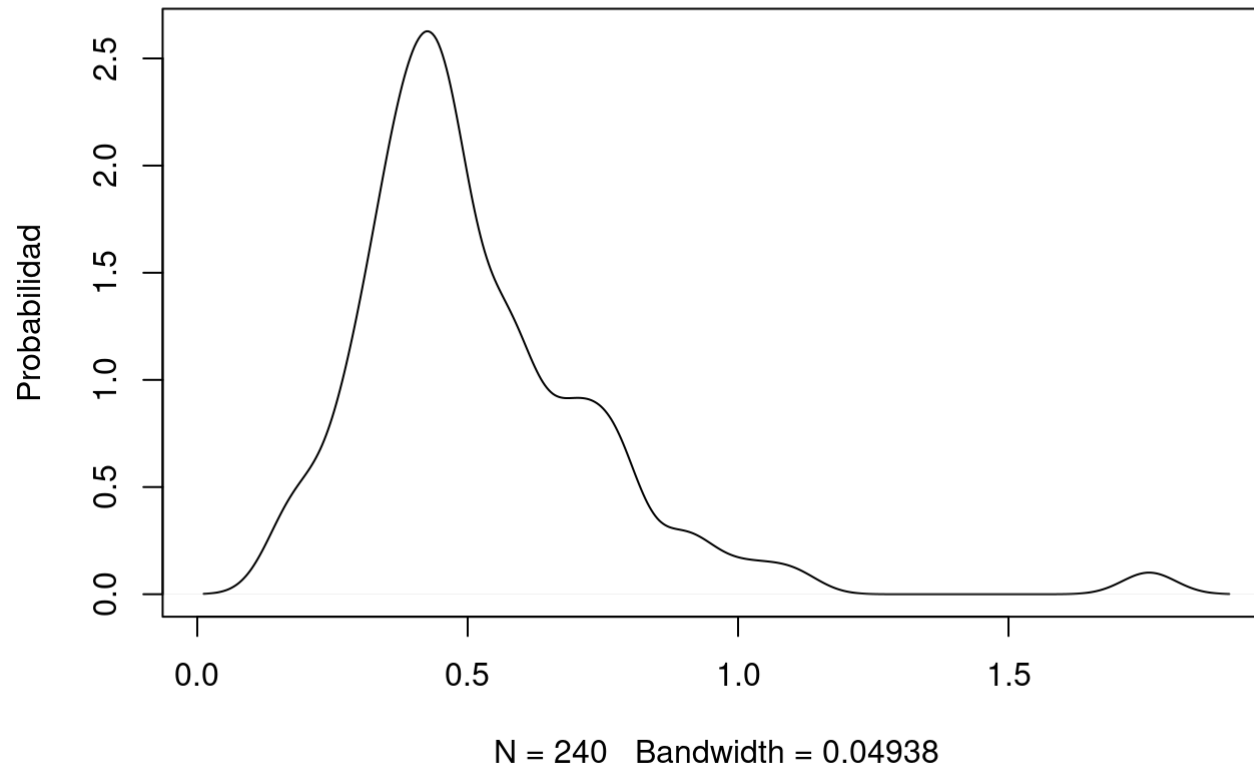
```
plot(density(Chlorella_mod$Temperatura), main = "Figura 4. Función de densidad para la variable Temperatura", y = "Probabilidad")
```

**Figura 4. Función de densidad para la variable Temperatura**



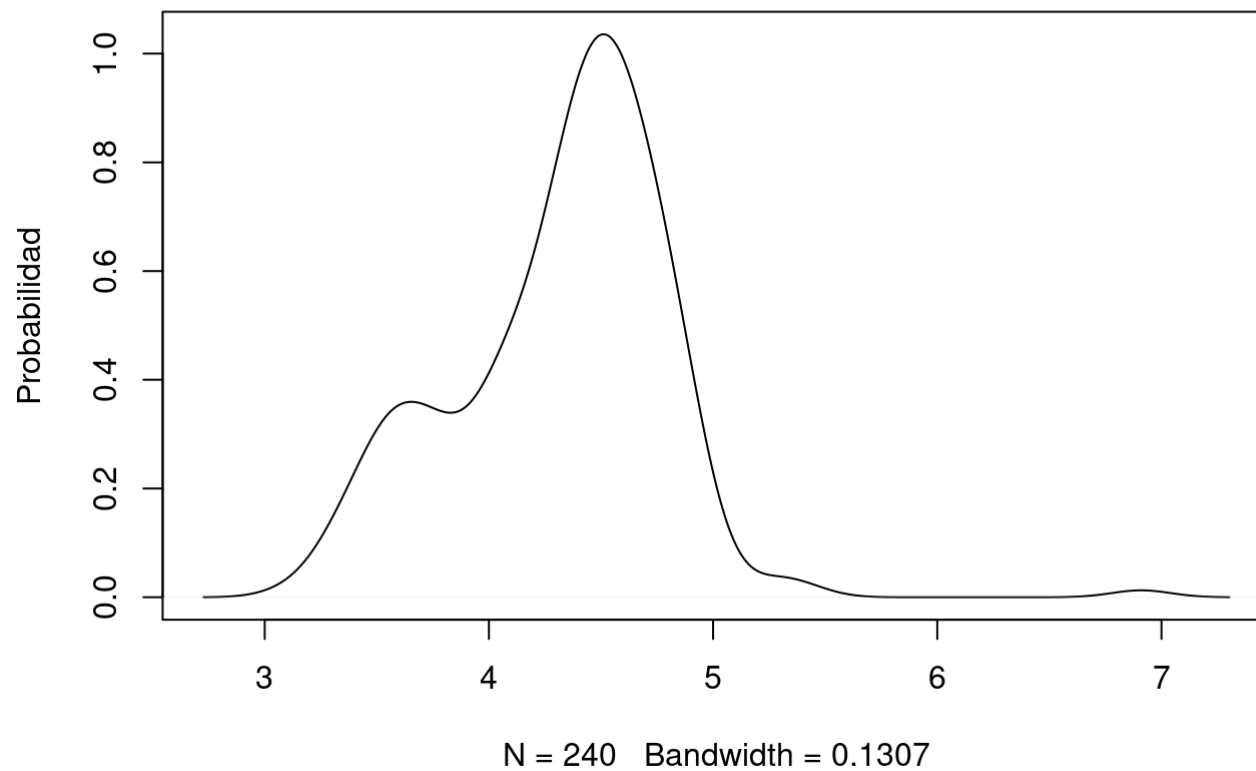
```
plot(density(Chlorella_mod$CO2), main = "Figura 5. Función de densidad para la variable CO2 (mg/ml)", y = "Probabilidad")
```

**Figura 5. Función de densidad para la variable CO2 (mg/ml)**



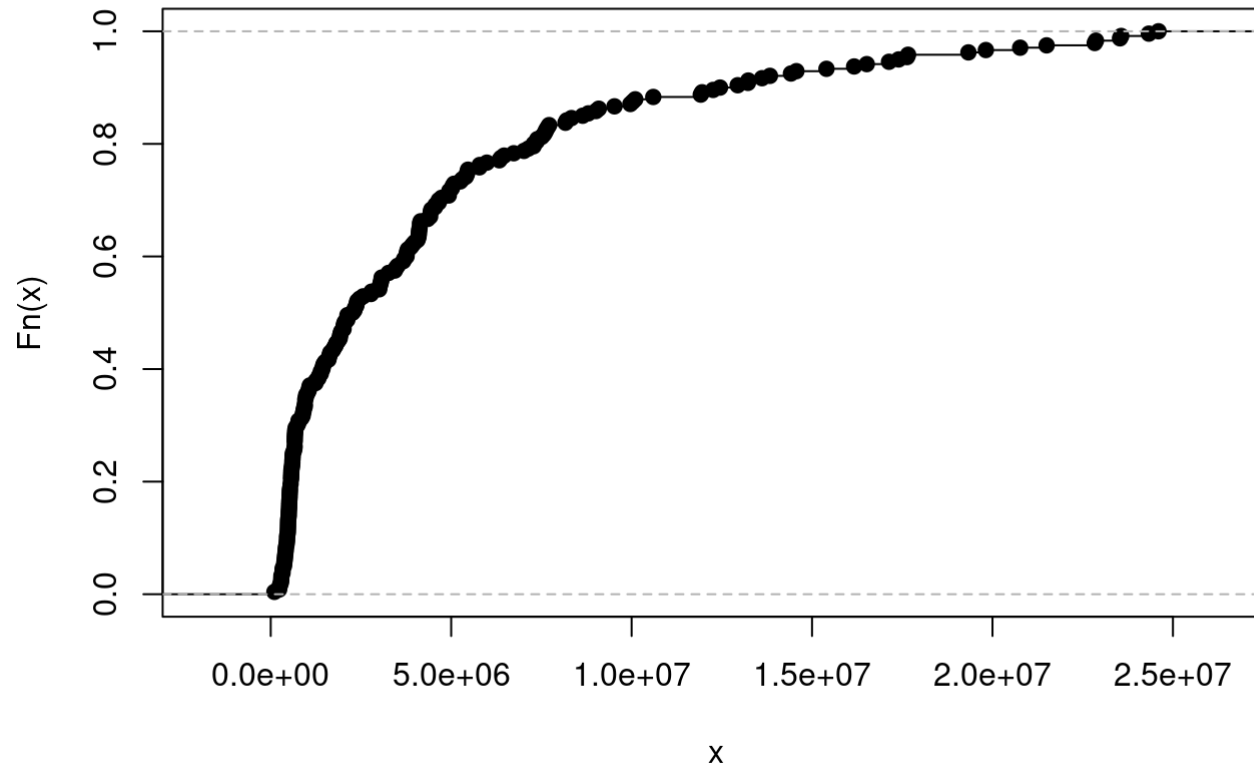
```
plot(density(Chlorella_mod$OD), main = "Figura 6. Función de densidad para la variable OD (mg/ml)", y = "Probabilidad")
```

**Figura 6. Función de densidad para la variable OD (mg/ml)**



```
plot(ecdf(Chlorella_mod$Densidad), main = "Figura 7. Función de probabilidad acumulada para la variable Densidad poblacional (cel/ml)")
```

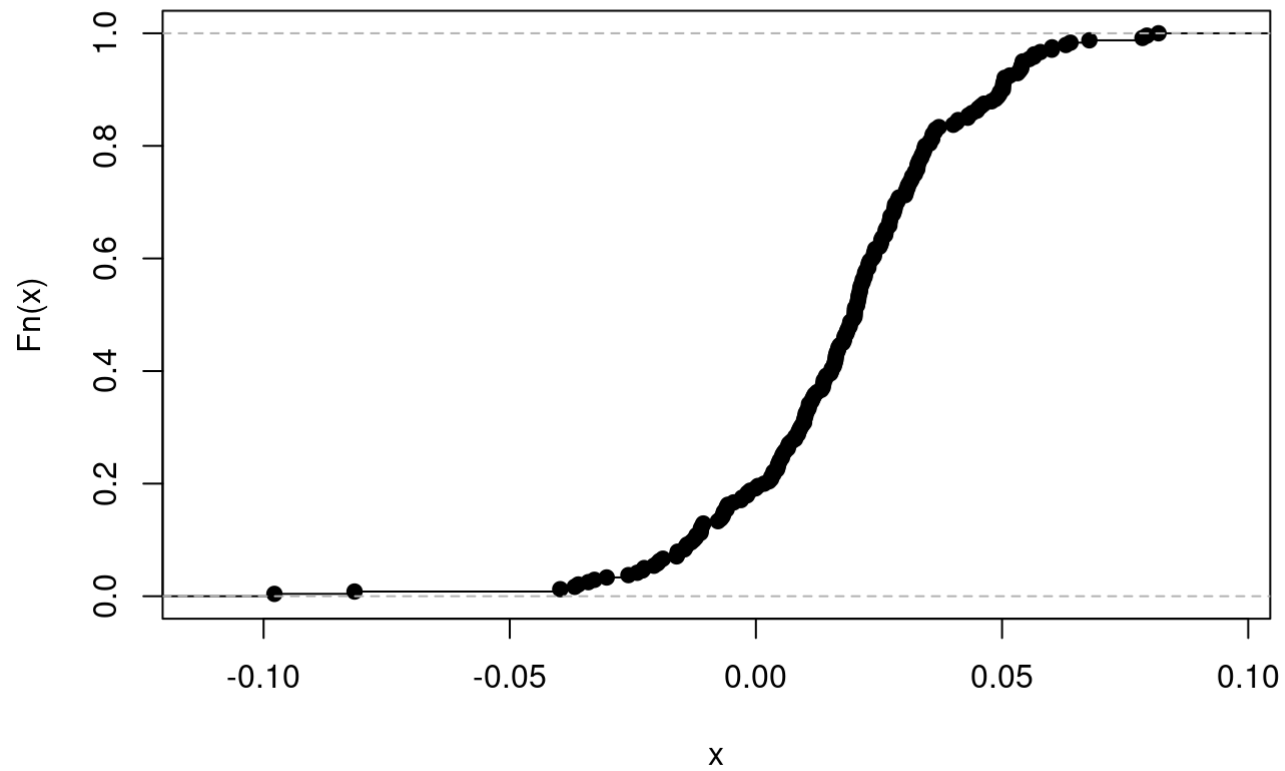
## 7. Función de probabilidad acumulada para la variable Densidad poblacio



```
plot(ecdf(Chlorella_mod$TCS), main = "Figura 8. Función de probabilidad acumulada para la variable TCS")
```

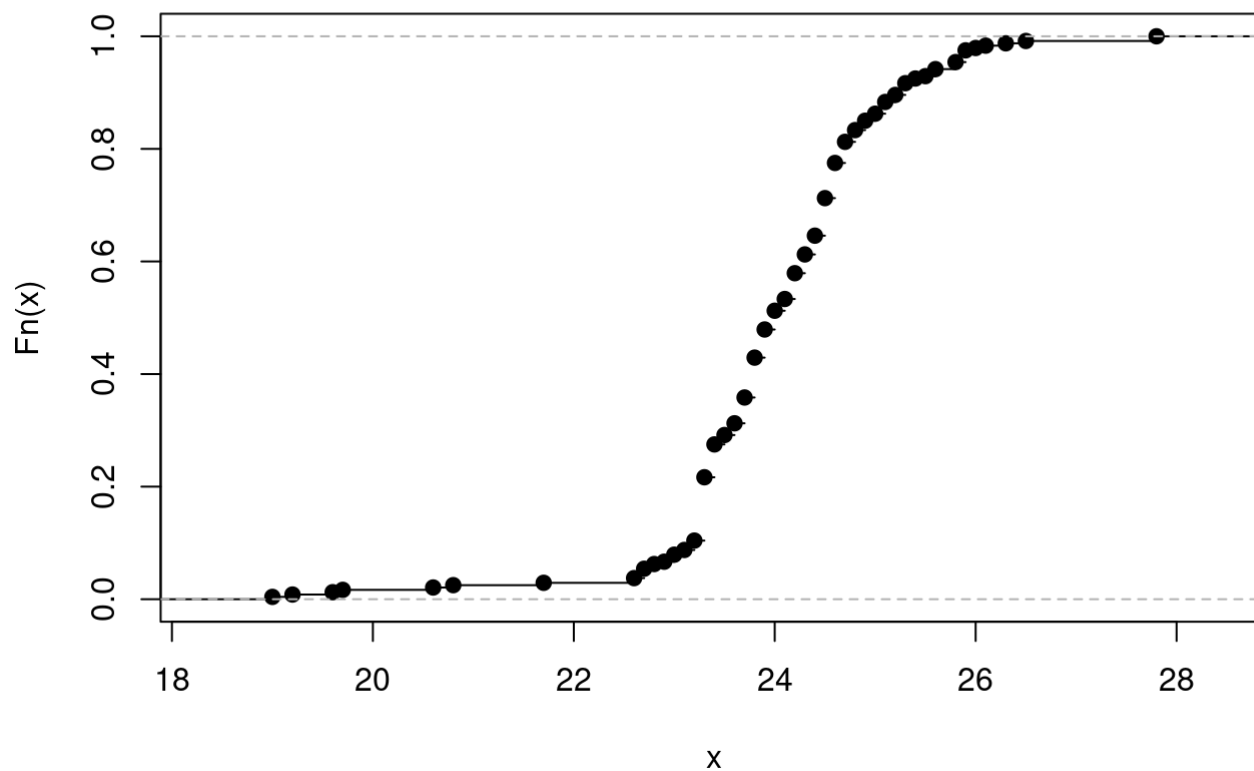


**Figura 8. Función de probabilidad acumulada para la variable TCS**



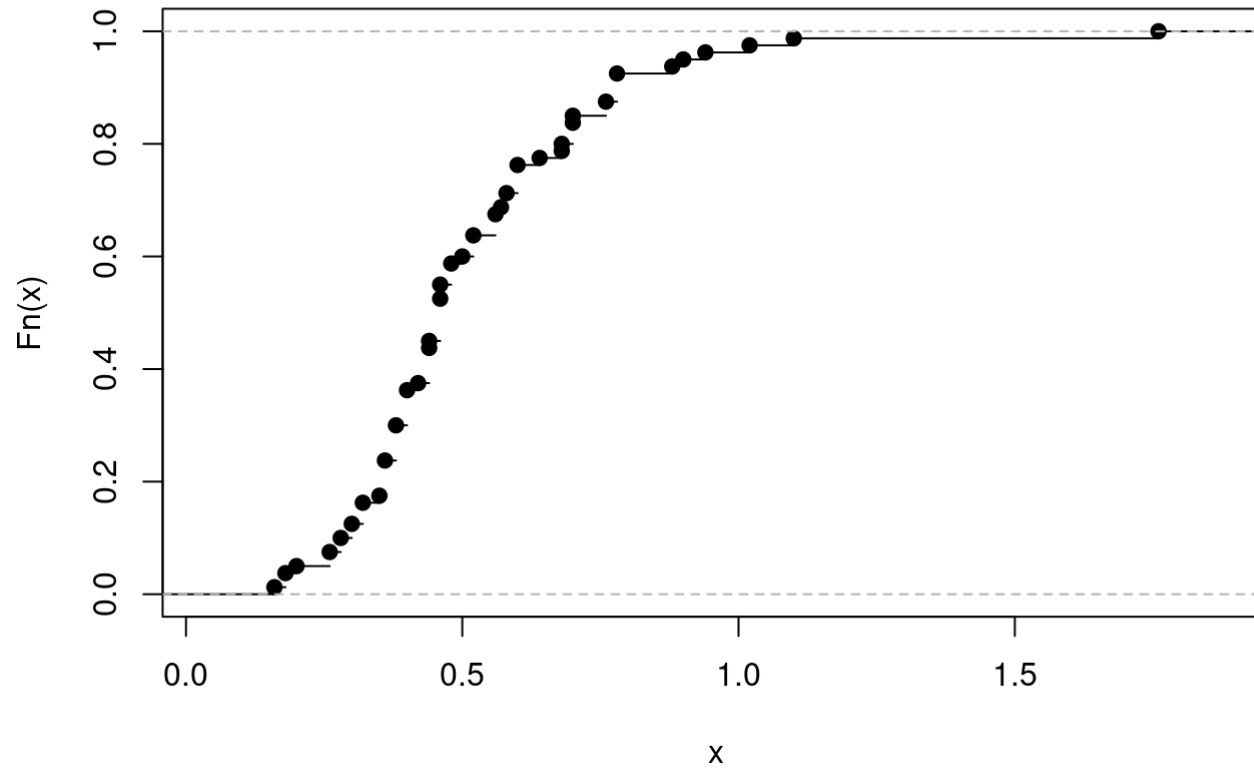
```
plot(ecdf(Chlorella_mod$Temperatura), main = "Figura 9. Función de probabilidad acumulada para la variable Temperatura")
```

**Figura 9. Función de probabilidad acumulada para la variable Temperatu**



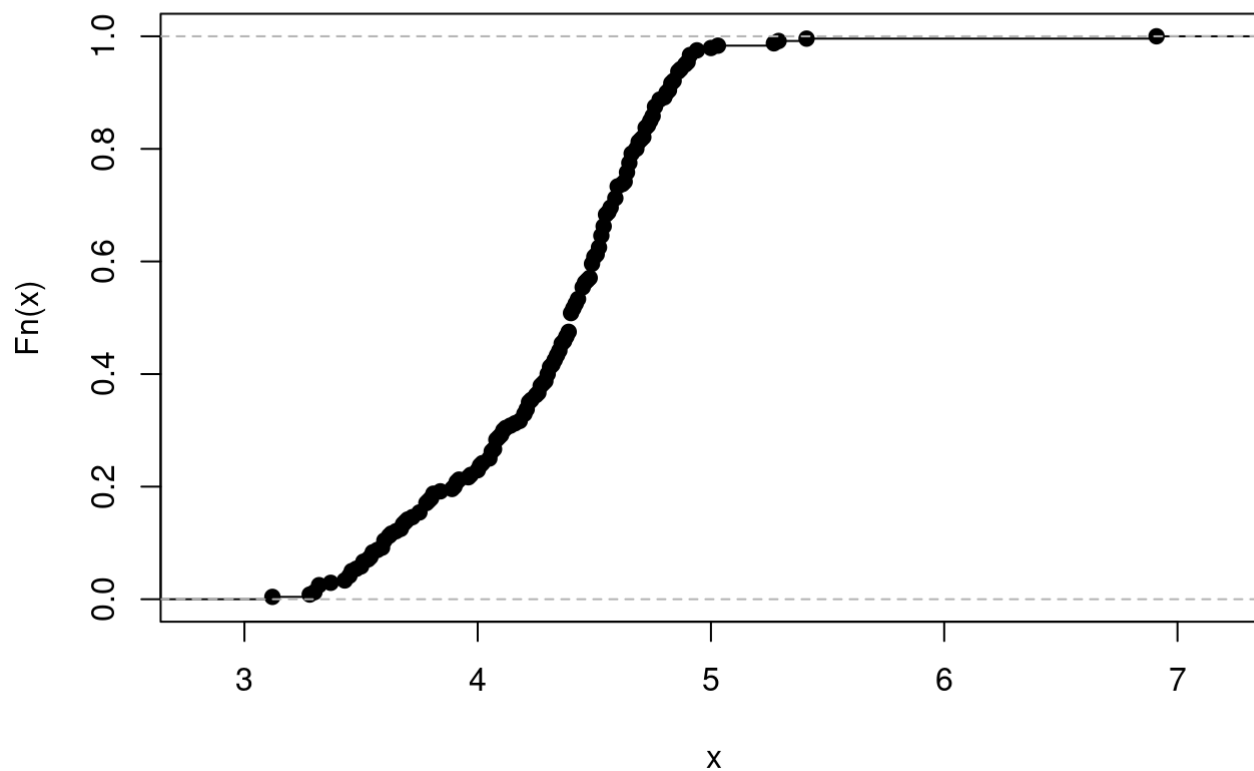
```
plot(ecdf(Chlorella_mod$C02), main = "Figura 10. Función de probabilidad acumulada para la variable C02 (mg/ml)")
```

**Figura 10. Función de probabilidad acumulada para la variable CO2 (mg/l**



```
plot(ecdf(Chlorella_mod$OD), main = "Figura 11. Función de probabilidad acumulada para la variable OD (mg/ml)")
```

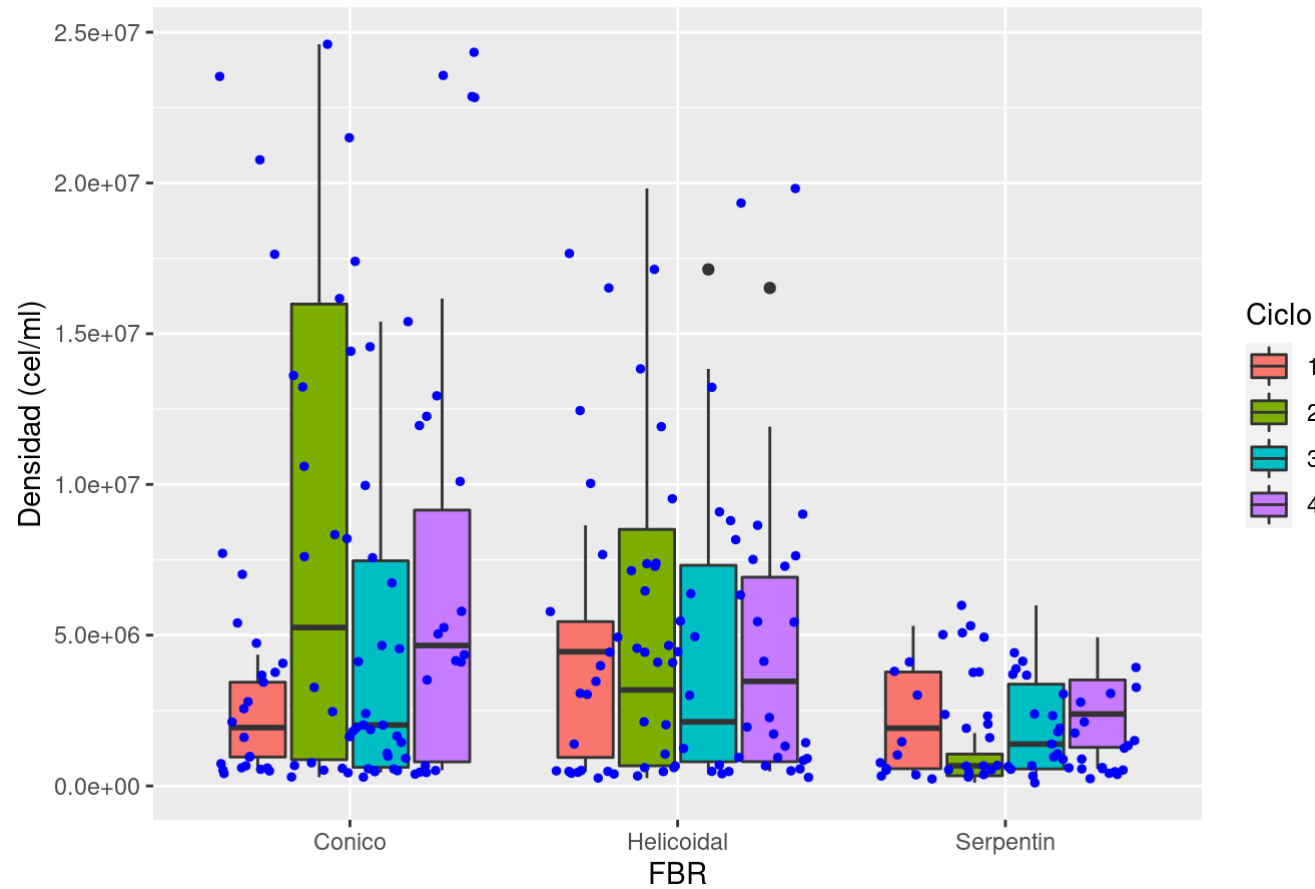
**Figura 11. Función de probabilidad acumulada para la variable OD (mg/n**



## Diagrama de cajas y bigotes

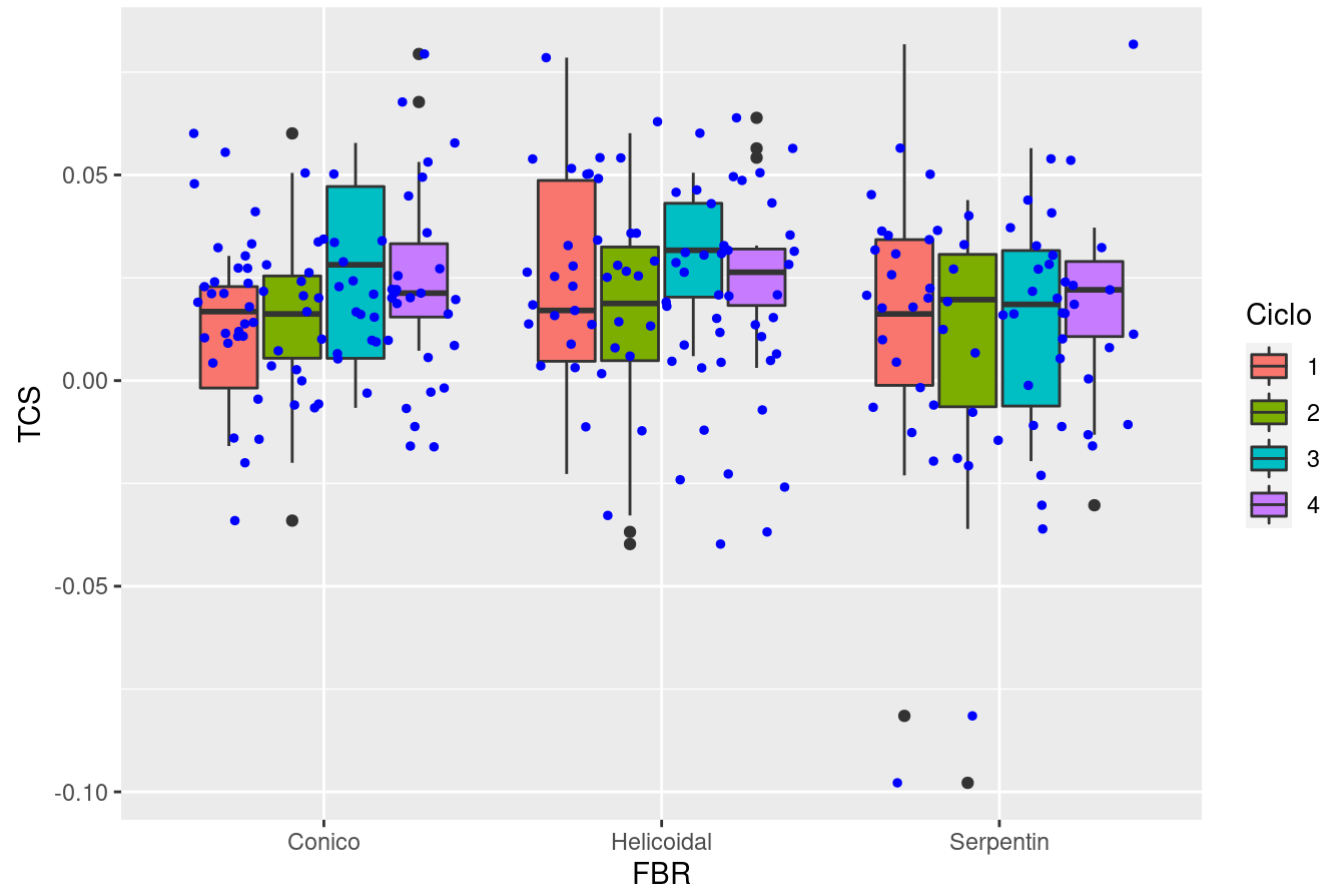
```
ggplot(Chlorella_mod, aes(x=FBR, y=Densidad)) +  
  geom_boxplot(aes(fill=Ciclo)) +  
  geom_jitter(color="blue", size=1, alpha=2) +  
  labs(title = "Figura 12. Densidad (cel/ml) agrupado por FBR y Ciclo", y="Densidad (cel/ml)")
```

Figura 12. Densidad (cel/ml) agrupado por FBR y Ciclo



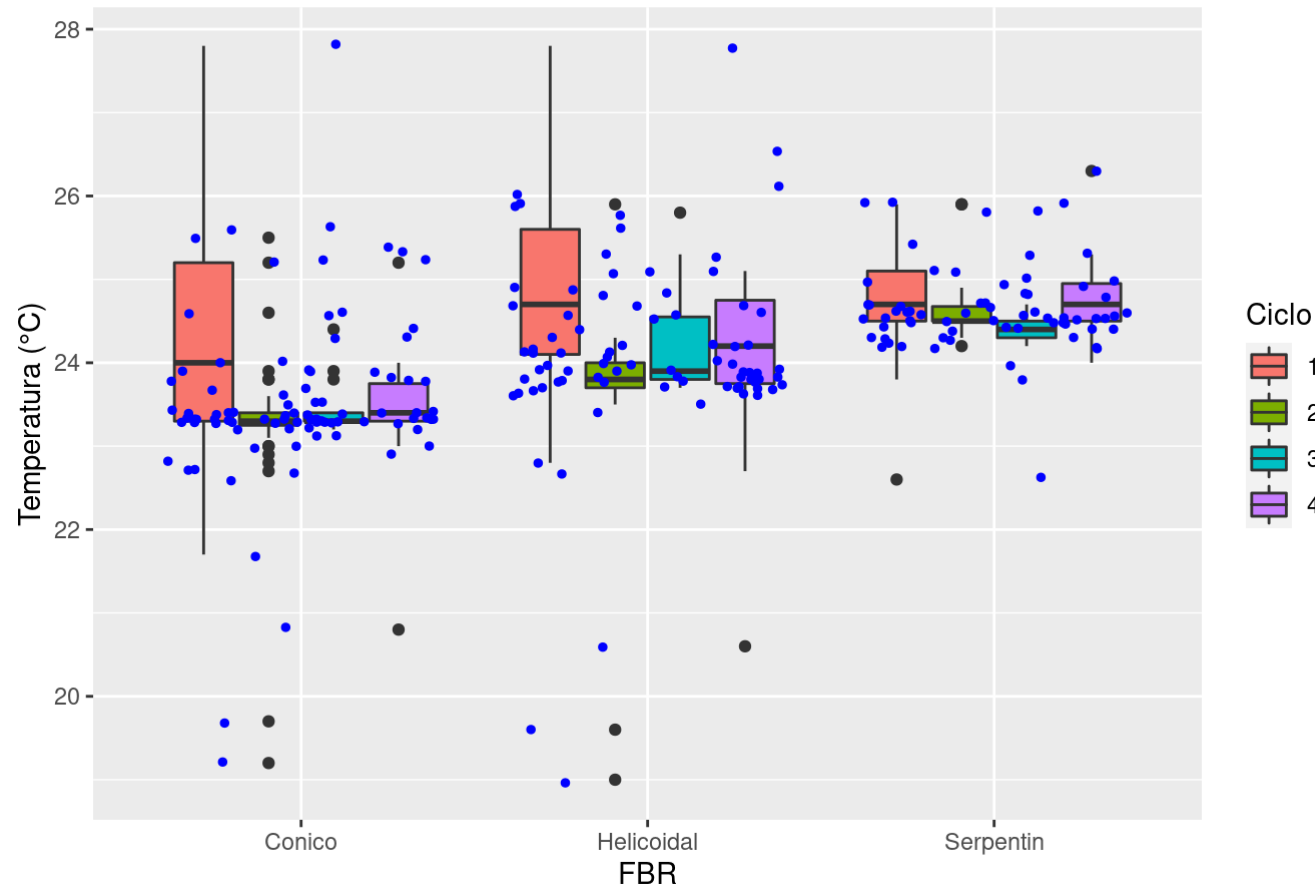
```
ggplot(Chlorella_mod, aes(x=FBR, y=TCS)) +
  geom_boxplot(aes(fill=Ciclo), na.rm=TRUE) +
  geom_jitter(color="blue", size=1, alpha=2) +
  labs(title = "Figura 13. Tasa de crecimiento simple (TCS) agrupado por FBR y Ciclo", y="TCS")
```

Figura 13. Tasa de crecimiento simple (TCS) agrupado por FBR y Ciclo



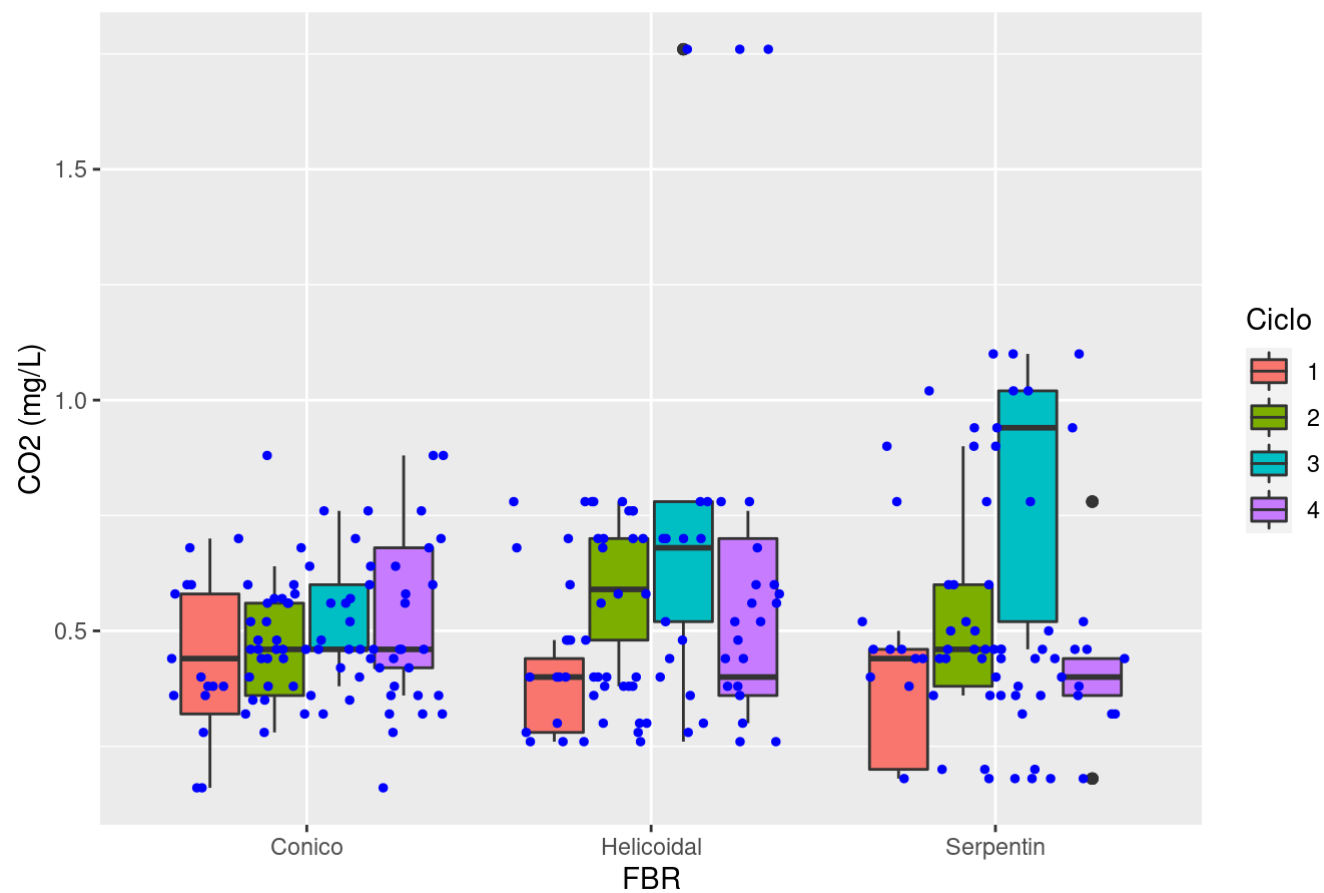
```
ggplot(Chlorella_mod, aes(x=FBR, y=Temperatura)) +
  geom_boxplot(aes(fill=Ciclo)) +
  geom_jitter(color="blue", size=1, alpha=2) +
  labs(title = "Figura 14. Temperatura (°C) agrupado por FBR y Ciclo", y="Temperatura (°C)")
```

Figura 14. Temperatura (°C) agrupado por FBR y Ciclo



```
ggplot(Chlorella_mod, aes(x=FBR, y=C02)) +  
  geom_boxplot(aes(fill=Ciclo)) +  
  geom_jitter(color="blue", size=1, alpha=2) +  
  labs(title = "Figura 15. C02 (mg/L) agrupado por FBR y Ciclo", y="C02 (mg/L)")
```

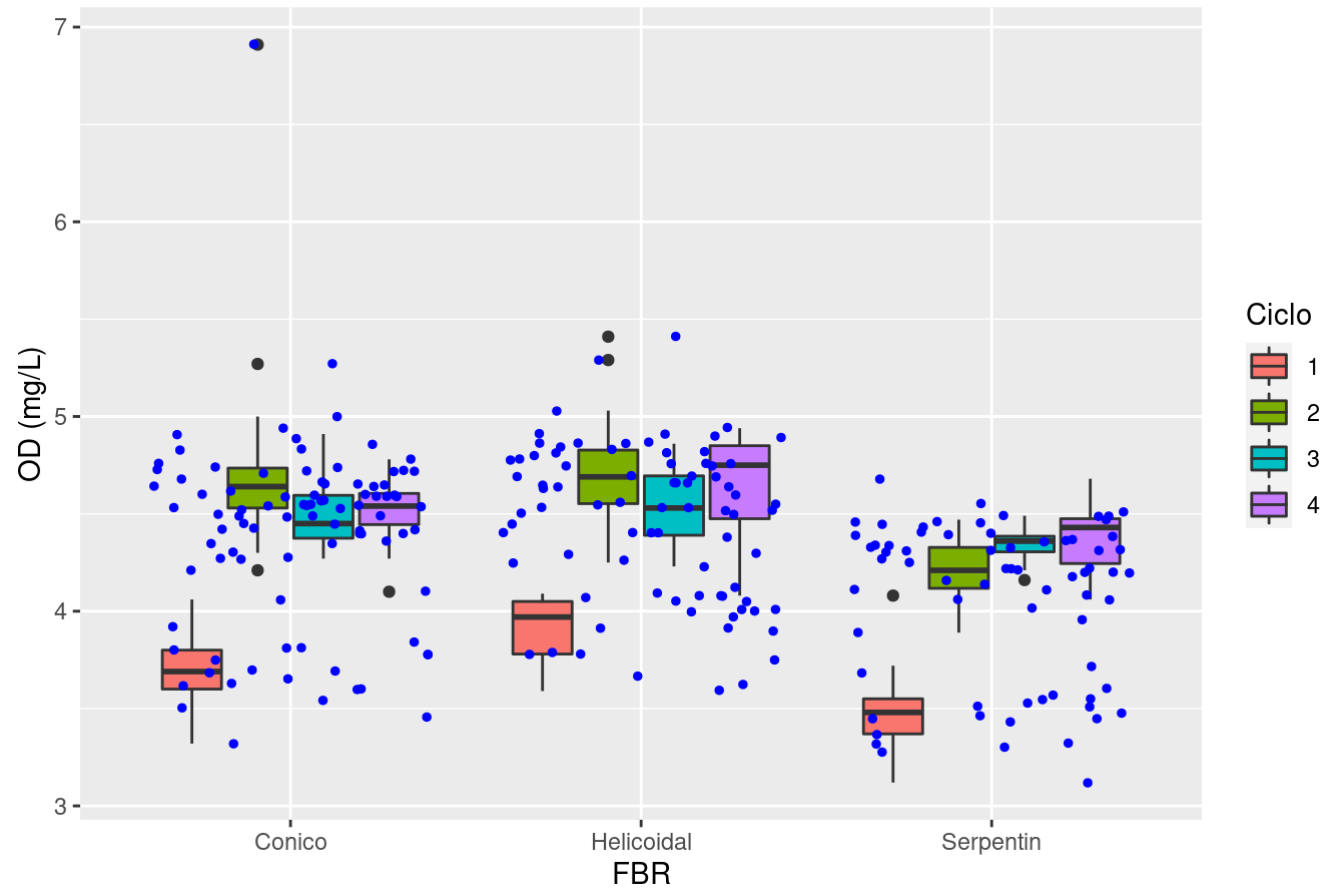
Figura 15. CO<sub>2</sub> (mg/L) agrupado por FBR y Ciclo



```
ggplot(Chlorella_mod, aes(x=FBR, y=OD)) +
  geom_boxplot(aes(fill=Ciclo)) +
  geom_jitter(color="blue", size=1, alpha=2) +
  labs(title = "Figura 16. Oxígeno disuelto (mg/L) agrupado por FBR y Ciclo", y="OD (mg/L)")
```



Figura 16. Oxígeno disuelto (mg/L) agrupado por FBR y Ciclo

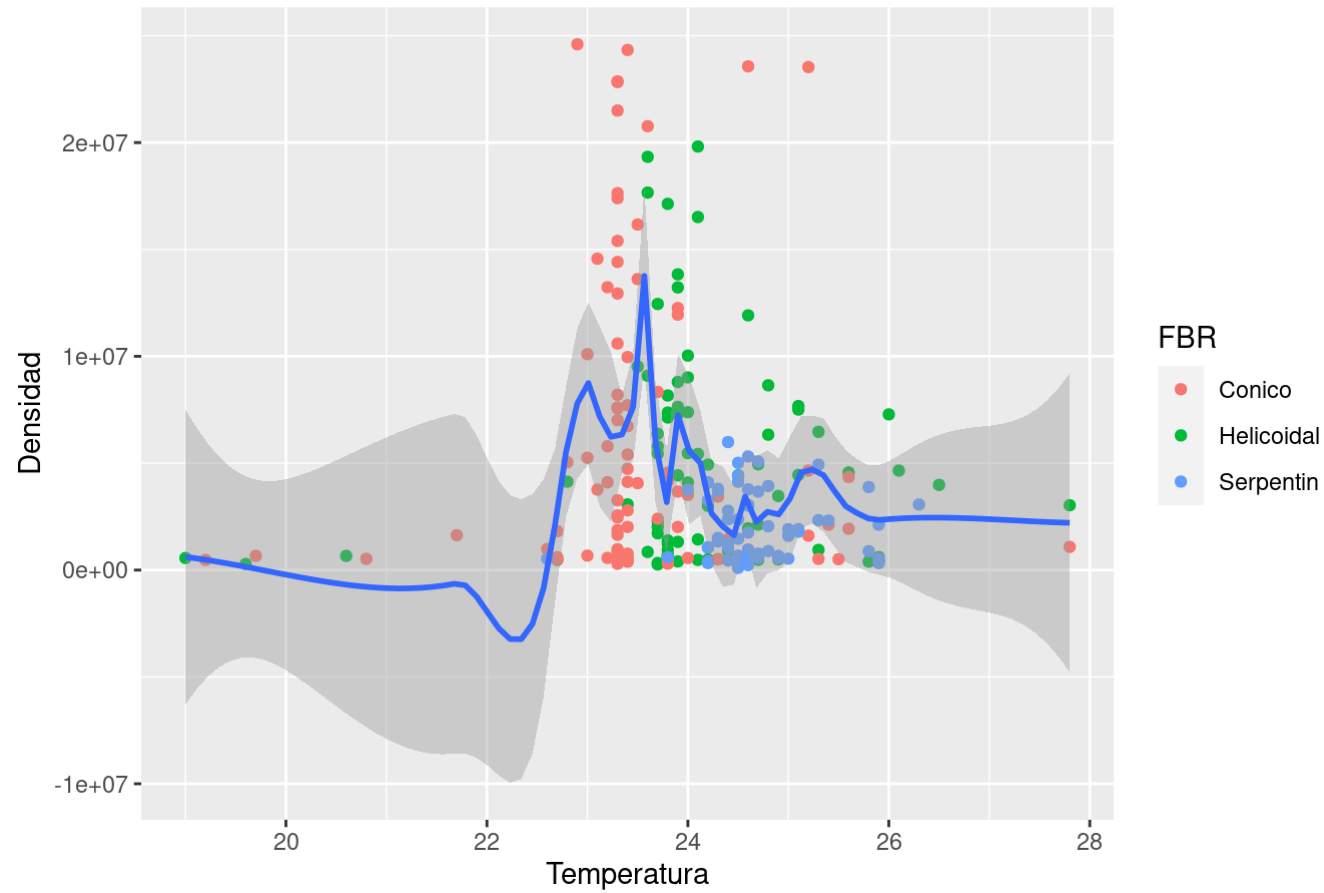


## Distribución y relación entre variables

```
ggplot(Chlorella_mod, aes(x=Temperatura, y=Densidad)) +  
  geom_point(aes(color=FBR)) +  
  geom_smooth(span=0.2) +  
  labs(title = "Figura 17. Relación entre Temperatura (°C) y Densidad (cel/ml) agrupadas por FBR")
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

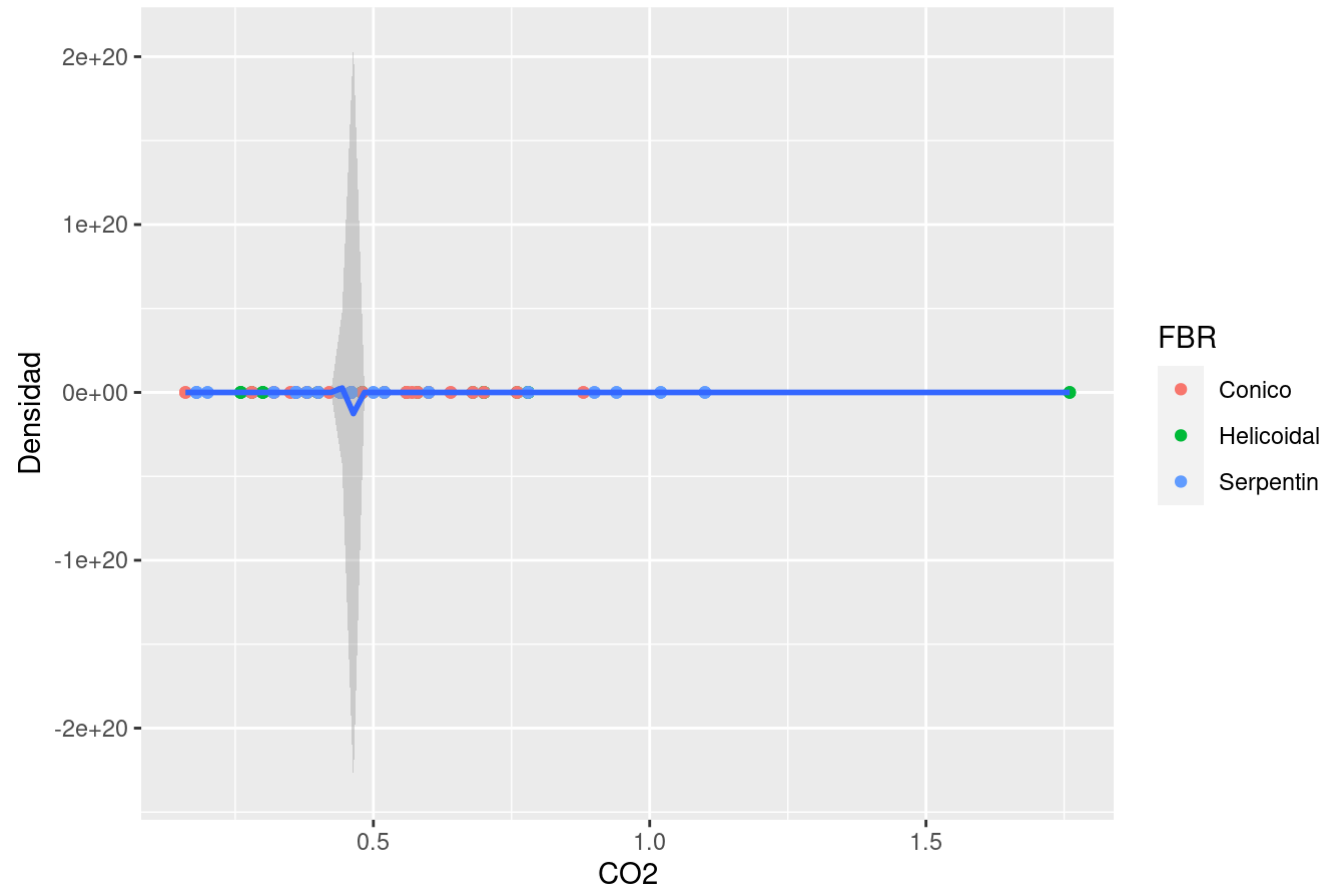
Figura 17. Relación entre Temperatura (°C) y Densidad (cel/ml) agrupadas p



```
ggplot(Chlorella_mod, aes(x=C02, y=Densidad)) +  
  geom_point(aes(color=FBR)) +  
  geom_smooth(span=0.2) +  
  labs(title = "Figura 18. Relación entre C02 (mg/L) y Densidad (cel/ml) agrupadas por FBR")
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

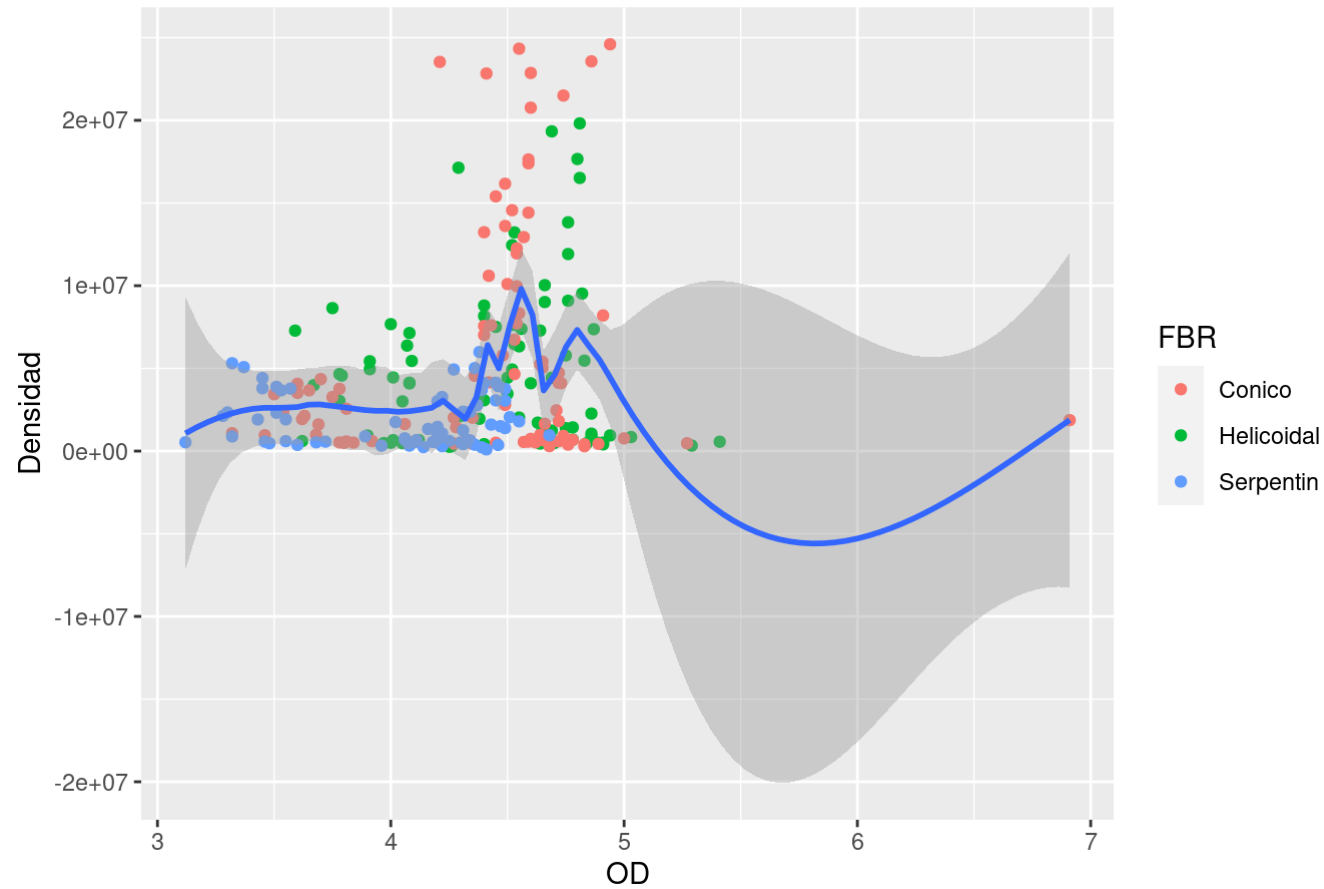
Figura 18. Relación entre CO2 (mg/L) y Densidad (cel/ml) agrupadas por FB



```
ggplot(Chlorella_mod, aes(x=OD, y=Densidad)) +
  geom_point(aes(color=FBR)) +
  geom_smooth(span=0.2) +
  labs(title = "Figura 19. Relación entre Oxígeno disuelto (mg/l) y Densidad (cel/ml) agrupadas por FBR")
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

Figura 19. Relación entre Oxígeno disuelto (mg/l) y Densidad (cel/ml) agrupada



## Correlación

```
Tabla_correlacion <- select(Chlorella_mod, Densidad, TCS, Temperatura, CO2, OD)
pairs.panels(Tabla_correlacion[,], smooth = TRUE, density = TRUE, method="pearson", font=1, main="Figura 20. Matriz de correlaciones")
```

**Figura 20. Matriz de correlaciones**

