Steatose

TP 0 : Start your first project !

What is metabolic hepatic steatosis ? How many mice are used? What treatments do they receive? How many groups are there?

- Hepatic steatosis is a common feature of metabolic liver diseases and an early marker of disease progression. Hepatic steatosis is characterized by an abnormal lipid accumulation within hepatocytes and may evolve with the appearance of hepatic lesions, such as hepatocyte ballooning and inflammation, that characterize the development of Metabolic Associated Steato-Hepatitis (MASH).

  On l'appelle aussi la maladie du soda je crois…

- Dans l'article, 97 souris âgées de 8 semaines sont utilisées pour l'étude. Elles sont réparties en 4 groupes.

Table 1: Répartition des groupes de souris utilisées

Groupe	Régime (Diet Control)	Injection
1	Yes	Iron-Dextran
2	Yes	Dextran only
3	No	Iron-Dextran
4	No	Dextran only

```
library(readr)
meta_data <- read_delim("Steatosis_phenotype.csv",
delim = ";", escape_double = FALSE, trim_ws = TRUE)
```

```
Rows: 96 Columns: 24
-- Column specification ----------------------------------------------------
```

```
Delimiter: ";"
chr  (2): Group, Type
dbl (22): ID, Steatosis, Fibrose, Hemolyse, Foci.surface, Biliruline tot, PA...

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
View(meta_data)
str(meta_data)
```

```
spc_tbl_ [96 x 24] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
 $ ID              : num [1:96] 2e+02 2e+03 2e+05 2e+06 2e+07 ...
 $ Group           : chr [1:96] "CON" "CON" "CON" "CON" ...
 $ Steatosis       : num [1:96] 0 0 0 0 0.1 0.05 0.15 0.8 0 0 ...
 $ Type            : chr [1:96] NA NA NA NA ...
 $ Fibrose         : num [1:96] 0 0 0 0 0 0 0 1 1 1 ...
 $ Hemolyse        : num [1:96] 0 0 0 0 0 0 0 0 0 0 ...
 $ Foci.surface    : num [1:96] 0.0844 0.1289 0.0321 0.1177 0.2223 ...
 $ Biliruline tot  : num [1:96] 0.9 1.2 2.4 1.2 0.3 0.6 0.6 0 2.4 3.6 ...
 $ PAL             : num [1:96] 39 42 57 51 33 24 36 45 66 45 ...
 $ ALAT            : num [1:96] 12 14.7 5.4 22.8 32.1 23.4 53.1 82.8 85.2 44.1 ...
 $ Cholesterol     : num [1:96] 1.86 2.61 1.74 2.76 3.18 2.55 3.93 4.14 2.85 2.79 ...
 $ Triglyceride    : num [1:96] 0.87 0.63 0.75 0.84 1.44 1.89 0.87 1.26 1.71 1.17 ...
 $ ASAT            : num [1:96] 92.1 68.4 48.3 114.6 93 ...
 $ serum.Iron      : num [1:96] 18.8 16.3 21.6 18.7 23.6 ...
 $ tf sat          : num [1:96] 40.8 38.1 49.3 42.9 29.9 ...
 $ AST/ALT         : num [1:96] 7.67 4.65 8.94 5.03 2.9 ...
 $ CK              : num [1:96] 381 378 117 942 432 168 243 90 525 639 ...
 $ Triglycerides Hp: num [1:96] 9.01 23 12.12 14.73 36.84 ...
 $ Hamp            : num [1:96] 6.07 6.69 6.65 6.6 4.17 ...
 $ Col1a           : num [1:96] -3.25 -4.67 -4.91 -5.03 -3.63 ...
 $ NRF2            : num [1:96] -0.059 -1.094 -0.551 -0.429 -0.797 ...
 $ A2M             : num [1:96] 7.04 6.7 6.29 5.91 6.77 ...
 $ Fsp27           : num [1:96] NA -7.96 -7.56 -8.15 -5.7 ...
 $ Crp             : num [1:96] 5.64 5.88 5.14 5.92 5.75 ...
 - attr(*, "spec")=
  .. cols(
  ..    ID = col_double(),
  ..    Group = col_character(),
  ..    Steatosis = col_double(),
  ..    Type = col_character(),
  ..    Fibrose = col_double(),
```

```
..    Hemolyse = col_double(),
..    Foci.surface = col_double(),
..    `Biliruline tot` = col_double(),
..    PAL = col_double(),
..    ALAT = col_double(),
..    Cholesterol = col_double(),
..    Triglyceride = col_double(),
..    ASAT = col_double(),
..    serum.Iron = col_double(),
..    `tf sat` = col_double(),
..    `AST/ALT` = col_double(),
..    CK = col_double(),
..    `Triglycerides Hp` = col_double(),
..    Hamp = col_double(),
..    Col1a = col_double(),
..    NRF2 = col_double(),
..    A2M = col_double(),
..    Fsp27 = col_double(),
..    Crp = col_double()
.. )
 - attr(*, "problems")=<externalptr>
```

Nous retirons les individus pour lesquels il manque des informations :

```
lines_with_na = which(is.na(apply(meta_data[,-c(2,4)], 1, sum)))
```

```
# Nombre total d'échantillons
total_samples <- nrow(meta_data)
cat("Nombre total de souris :", total_samples, "\n")
```

Nombre total de souris : 96

```
# Nombre de groupes de traitement
nb_groups <- length(unique(meta_data$Group))
cat("Nombre de groupes de traitement :", nb_groups, "\n")
```

Nombre de groupes de traitement : 4

```
samples_per_group <- table(meta_data$Group)
print("Nombre de souris par groupe :")
```

```
[1] "Nombre de souris par groupe :"
```

```
print(samples_per_group)
```

```
    CON      HF HF+IRON    IRON
     23      23      25      25
```

Il y a une souris en moins qu'annoncées. Cela doit être expliqué dans l'article, mais il faudrait retrouver la citation.

```
library(dplyr)
```

```
Attaching package: 'dplyr'
```

```
The following objects are masked from 'package:stats':

    filter, lag
```

```
The following objects are masked from 'package:base':

    intersect, setdiff, setequal, union
```

```
meta_data %>%
group_by(Group) %>%
summarise(mean = mean(Cholesterol,na.rm = TRUE),
std = sd(Cholesterol,na.rm = TRUE),
n = n())
```

```
# A tibble: 4 x 4
  Group    mean   std     n
  <chr>   <dbl> <dbl> <int>
1 CON      2.51 0.579    23
2 HF       3.82 0.766    23
3 HF+IRON  3.74 0.553    25
4 IRON     2.43 0.232    25
```

Nous faisons les boxplots :

Avec \ggplot2 :

```r
library(ggplot2)

# Avec couleurs et thème
ggplot(meta_data, aes(x = Group, y = Cholesterol, fill = Group)) +
  geom_boxplot() +
  labs(title = "Boxplot du cholestérol des souris en fonction de leur groupe",
       x = "Groupes",
       y = "Teneur en cholestérol") +
  theme_minimal() +
  theme(legend.position = "none")
```

```
Warning: Removed 8 rows containing non-finite outside the scale range
(`stat_boxplot()`).
```
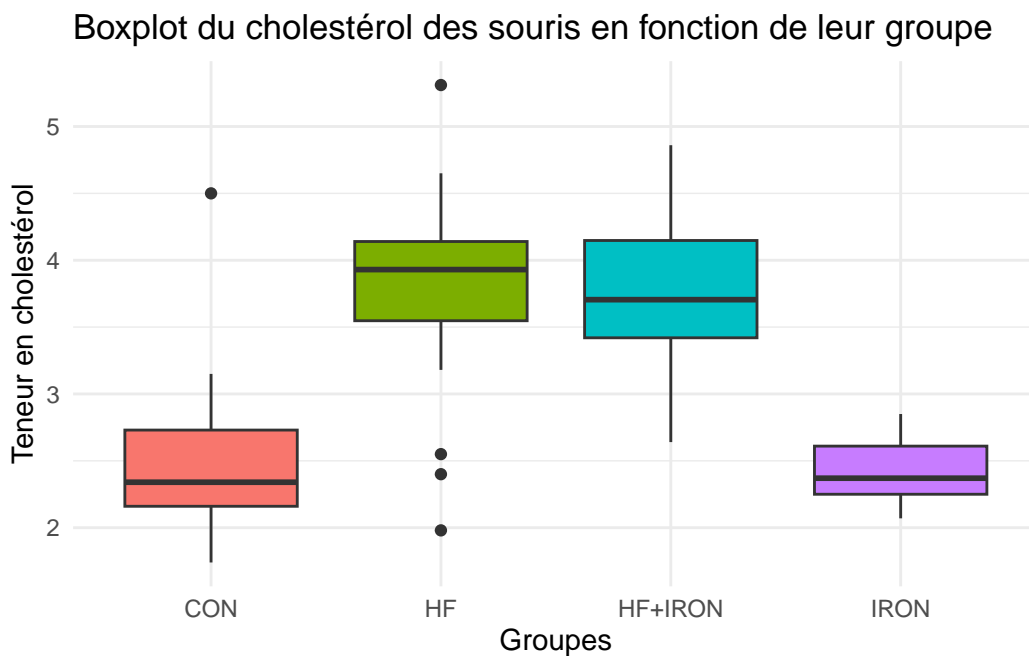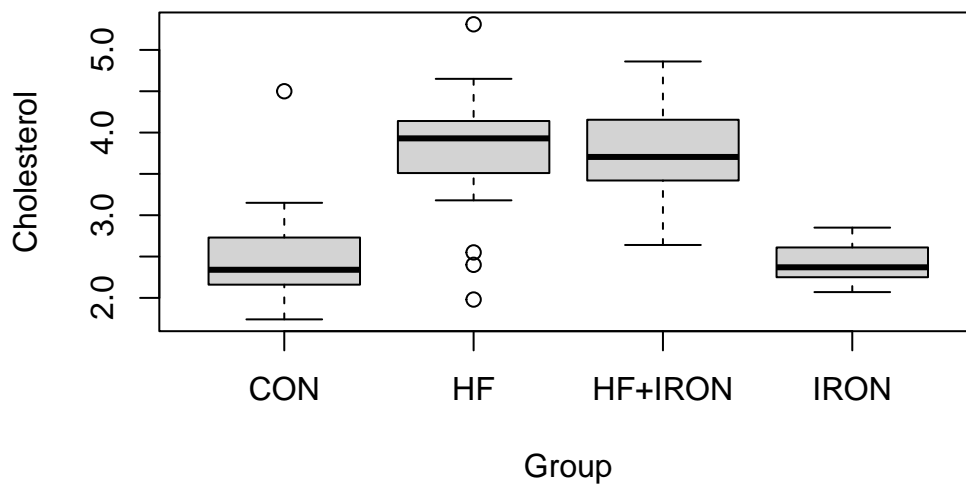

Sans ggplot2 (test) :

```r
# boxplot pour l'entièreté du jeu de données
boxplot(meta_data$Cholesterol)
```
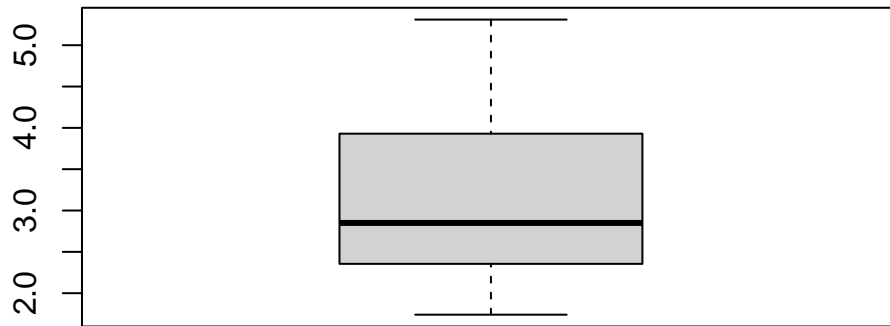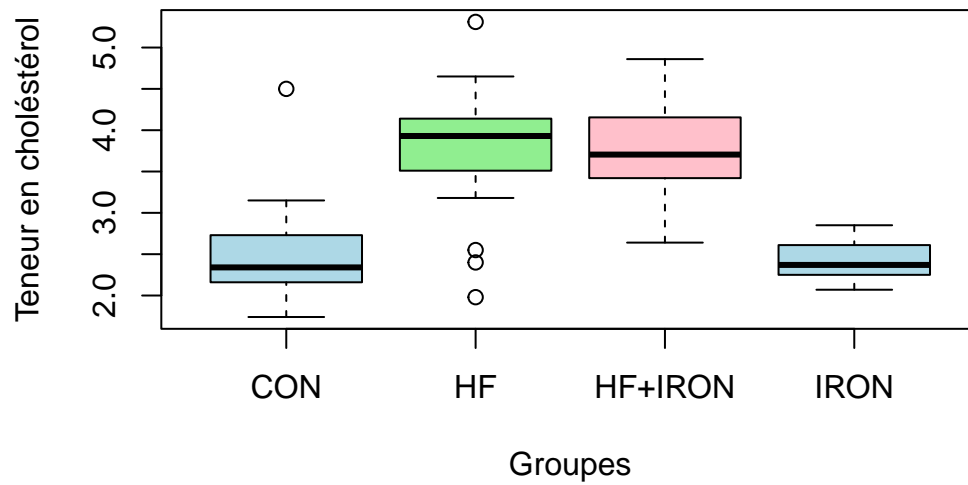
```
# Boxplot par groupe
boxplot(Cholesterol ~ Group, data = meta_data)
```



```
# avec des couleurs
boxplot(Cholesterol ~ Group,
        data = meta_data,
        main = "Boxplot du choléstérol des souris en fonction de leur groupe",
        xlab = "Groupes",
        ylab = "Teneur en choléstérol",
        col = c("lightblue", "lightgreen", "pink"))
```

Boxplot du choléstérol des souris en fonction de leur grou


```
library(tidyr) #rajout de la librarie nécessaire

df.explore = meta_data[,-c(1,4)]
resume_stats <- df.explore %>%
pivot_longer(cols = -Group,
names_to = "Variable",
values_to = "Valeur") %>%
group_by(Group, Variable) %>%
summarise(
Mean = mean(Valeur, na.rm = TRUE),
SD = sd(Valeur, na.rm = TRUE),
.groups = "drop"
)
```