

Document partagé "Introduction au logiciel R - 2024"

Mercredi matin : rappels

Vincent : RStudio 2023.06.1+524 "Mountain Hydrangea" Release
(547dcf861cac0253a8abb52c135e44e02ba407a1, 2023-07-07) for windows
Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko)
RStudio/2023.06.1+524 Chrome/110.0.5481.208 Electron/23.3.0 Safari/537.36

RStudio 2023.12.1+402 "Ocean Storm" Release
(4da58325ffcff29d157d9264087d4b1ab27f7204, 2024-01-28) for windows
Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko)
RStudio/2023.12.1+402 Chrome/116.0.5845.190 Electron/26.2.4 Safari/537.36

RStudio 2023.12.1+402 "Ocean Storm" Release
(4da58325ffcff29d157d9264087d4b1ab27f7204, 2024-01-28) for windows
Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko)
RStudio/2023.12.1+402 Chrome/116.0.5845.190 Electron/26.2.4 Safari/537.36

Inès :

RStudio 2023.06.1+524 "Mountain Hydrangea" Release
(547dcf861cac0253a8abb52c135e44e02ba407a1, 2023-07-07) for windows
Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko)
RStudio/2023.06.1+524 Chrome/110.0.5481.208 Electron/23.3.0 Safari/537.36

Les fonctions utiles :

```
table(fruits$groupe)
barplot(table(fruits$groupe))
```

Carole

RStudio 2023.06.1+524 "Mountain Hydrangea" Release
(547dcf861cac0253a8abb52c135e44e02ba407a1, 2023-07-07) for windows
Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko)
RStudio/2023.06.1+524 Chrome/110.0.5481.208 Electron/23.3.0 Safari/537.36

Agnes

RStudio 2023.12.1+402 "Ocean Storm" Release
(4da58325ffcff29d157d9264087d4b1ab27f7204, 2024-01-28) for windows
Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko)
RStudio/2023.12.1+402 Chrome/116.0.5845.190 Electron/26.2.4 Safari/537.36

paul

RStudio 2023.12.1+402 "Ocean Storm" Release
(4da58325ffcff29d157d9264087d4b1ab27f7204, 2024-01-28) for windows
Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko)
RStudio/2023.12.1+402 Chrome/116.0.5845.190 Electron/26.2.4 Safari/537.36

Isabelle

RStudio 2023.12.1+402 "Ocean Storm" Release
(4da58325ffcff29d157d9264087d4b1ab27f7204, 2024-01-28) for windows
Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko)
RStudio/2023.12.1+402 Chrome/116.0.5845.190 Electron/26.2.4 Safari/537.36

RStudio 2023.06.1+524 "Mountain Hydrangea" Release
(547dcf861cac0253a8abb52c135e44e02ba407a1, 2023-07-07) for windows
Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like

Les fonctions utiles :

```
table(fruits$groupe)
barplot(table(fruits$groupe))
```

Gecko) RStudio/2023.06.1+524 Chrome/110.0.5481.208 Electron/23.3.0 Safari/537.36

Emilie

RStudio 2023.12.1+402 "Ocean Storm" Release

(4da58325ffcff29d157d9264087d4b1ab27f7204, 2024-01-28) for windows

Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko)

RStudio/2023.12.1+402 Chrome/116.0.5845.190 Electron/26.2.4 Safari/537.36

RStudio 2023.12.1+402 "Ocean Storm" Release

(4da58325ffcff29d157d9264087d4b1ab27f7204, 2024-01-29) for macOS

Mozilla/5.0 (Macintosh; Intel Mac OS X 10_15_7) AppleWebKit/537.36 (KHTML, like Gecko)

RStudio/2023.12.1+402 Chrome/116.0.5845.190 Electron/26.2.4 Safari/537.36

[illegible]

Mahdi

RStudio 2023.06.1+524 "Mountain Hydrangea" Release

(547dcf861cac0253a8abb52c135e44e02ba407a1, 2023-07-07) for windows

Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko)

RStudio/2023.06.1+524 Chrome/110.0.5481.208 Electron/23.3.0 Safari/537.36

Corentin :

RStudio 2023.09.1+494 "Desert Sunflower" Release

(cd7011dce393115d3a7c3db799dda4b1c7e88711, 2023-10-16) for windows

Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko)

RStudio/2023.09.1+494 Chrome/116.0.5845.190 Electron/26.2.4 Safari/537.36

Charlotte:

Les fonctions utiles :

```
table(fruits$groupe)
```

```
barplot(table(fruits$groupe))
```

RStudio 2022.07.1+554 "Spotted Wakerobin" Release
(7872775ebddc40635780ca1ed238934c3345c5de, 2022-07-22) for macOS
Mozilla/5.0 (Macintosh; Intel Mac OS X 10_15_7) AppleWebKit/537.36 (KHTML, like Gecko)
QtWebEngine/5.12.10 Chrome/69.0.3497.128 Safari/537.36

Manon
RStudio 2023.12.1+402 "Ocean Storm" Release
(4da58325ffcff29d157d9264087d4b1ab27f7204, 2024-01-29) for macOS
Mozilla/5.0 (Macintosh; Intel Mac OS X 10_15_7) AppleWebKit/537.36 (KHTML, like Gecko)
RStudio/2023.12.1+402 Chrome/116.0.5845.190 Electron/26.2.4 Safari/537.36

Anissa

RStudio 2023.12.1+402 "Ocean Storm" Release
(4da58325ffcff29d157d9264087d4b1ab27f7204, 2024-01-28) for windows
Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko)
RStudio/2023.12.1+402 Chrome/116.0.5845.190 Electron/26.2.4 Safari/537.36

RStudio

2023.12.1 Build 402

© 2009-2024 Posit Software, PBC

"Ocean Storm" Release (4da58325, 2024-01-28) for windows
Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) RStudio/2023.12.1+402 Chrome/116.0.5845.190
Electron/26.2.4 Safari/537.36

RStudio

2023.06.1 Build 524

© 2009-2023 Posit Software, PBC

Les fonctions utiles :

```
table(fruits$groupe)
barplot(table(fruits$groupe))
```

LAN

Thomas

RStudio 2023.12.1+402 "Ocean Storm" Release
(4da58325ffcff29d157d9264087d4b1ab27f7204, 2024-01-29) for macOS
Mozilla/5.0 (Macintosh; Intel Mac OS X 10_15_7) AppleWebKit/537.36 (KHTML, like Gecko)
RStudio/2023.12.1+402 Chrome/116.0.5845.190 Electron/26.2.4 Safari/537.36

Maité
RStudio 2021.09.2+382 "Ghost Orchid" Release
(fc9e217980ee9320126e33cdf334d4f4e105dc4f, 2022-01-04) for macOS
Mozilla/5.0 (Macintosh; Intel Mac OS X 13_6_1) AppleWebKit/537.36 (KHTML, like Gecko) QtWebEngine/5.12.10 Chrome/69.0.3497.128 Safari/537.36

Idris :

RStudio

2023.12.1 Build 402

© 2009-2024 Posit Software, PBC

"Ocean Storm" Release (4da58325, 2024-01-29) for macOS
Mozilla/5.0 (Macintosh; Intel Mac OS X 10_15_7) AppleWebKit/537.36 (KHTML, like Gecko)
RStudio/2023.12.1+402 Chrome/116.0.5845.190 Electron/26.2.4 Safari/537.36

Les fonctions utiles :

```
table(fruits$groupe)
barplot(table(fruits$groupe))
```

Camille

2022.12.0 Build 353

Bastien et Marie:

RStudio 2023.12.1 Build 402

© 2009-2024 Posit Software, PBC

"Ocean Storm" Release (4da58325, 2024-01-28) for windows

Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko)

RStudio/2023.12.1+402 Chrome/116.0.5845.190Electron/26.2.4 Safari/537.36

Asma

RStudio

2023.12.1 Build 402

© 2009-2024 Posit Software, PBC

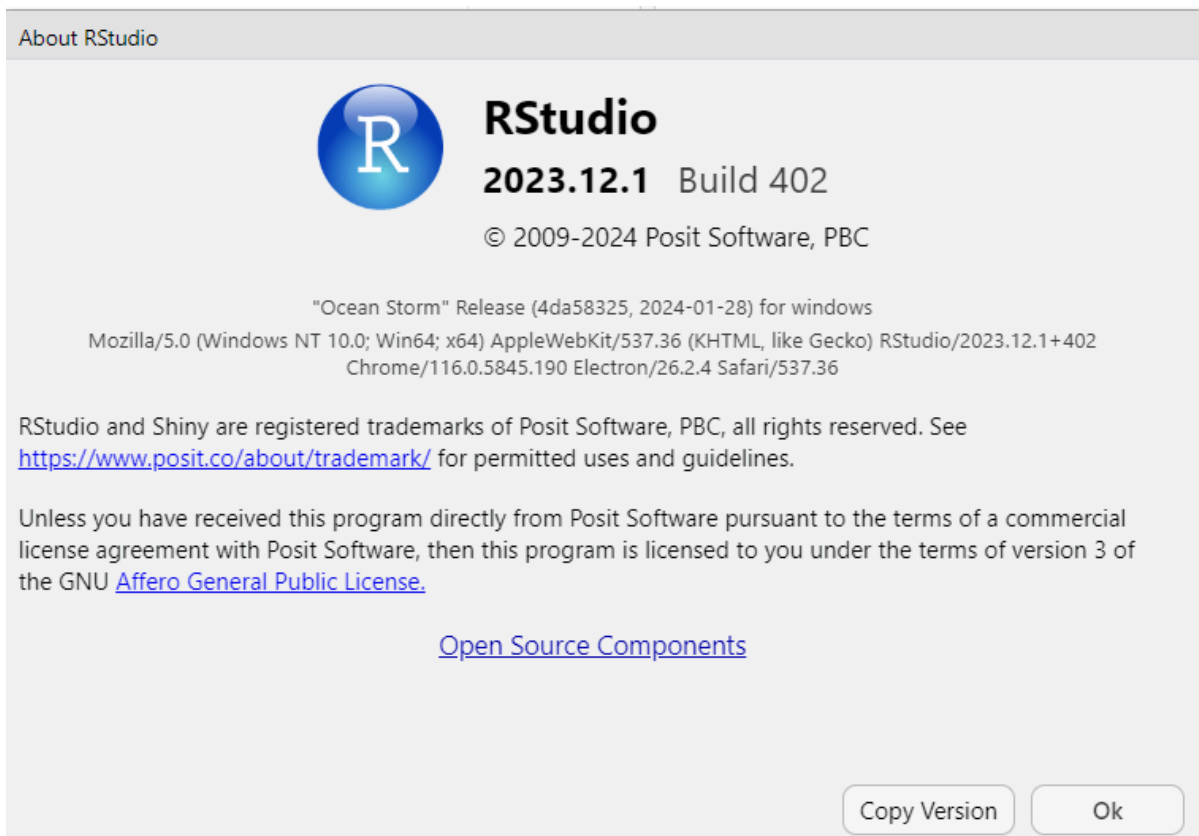
"Ocean Storm" Release (4da58325, 2024-01-28) for windows

Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) RStudio/2023.12.1+402 Chrome/116.0.5845.190
Electron/26.2.4 Safari/537.36

Les fonctions utiles :

```
table(fruits$groupe)
```

```
barplot(table(fruits$groupe))
```



Exercice installation du package intro2r

```
remotes::install_github("vguillemot/intro2r")
```

Les fonctions utiles :

```
table(fruits$groupe)  
barplot(table(fruits$groupe))
```

```
Les packages binaires téléchargés sont dans
C:\Users\Libre\AppData\Local\Temp\RtmpGE8TGe\downloaded_packages
> remotes::install_github("vguillemot/intro2r")
Error in loadNamespace(x) : aucun package nommé 'remotes' n'est trouvé
```

Deja installé?

Skipping install of 'intro2r' from a github remote, the SHA1 (deff5d41) has not changed since last install.
Use `force = TRUE` to force installation

```
Warning in install.packages :
  impossible d'accéder à l'index de l'entrepôt
https://cran.rstudio.com/src/contrib:
  impossible d'ouvrir l'URL 'https://cran.rstudio.com/src/contrib/PACKAGES'
Warning in install.packages :
  le package 'remote' n'est pas disponible for this version of R
```

Une version de ce package pour votre version de R est peut-être disponible ailleurs,
Voyez des idées à
<https://cran.r-project.org/doc/manuals/r-patched/R-admin.html#Installing-packages>

```
Warning in install.packages :
  impossible d'accéder à l'index de l'entrepôt
https://cran.rstudio.com/bin/windows/contrib/4.3:
  impossible d'ouvrir l'URL
'https://cran.rstudio.com/bin/windows/contrib/4.3/PACKAGES'
```

Warning message:
package 'remotes' was built under R version 4.2.3

Les fonctions utiles :

```
table(fruits$groupe)
barplot(table(fruits$groupe))
```



```
> remotes::install_github("vguillemot/intro2r")
Downloading GitHub repo vguillemot/intro2r@HEAD
Error in utils::download.file(url, path, method = method, quiet = quiet, :
  le téléchargement de
'https://api.github.com/repos/vguillemot/intro2r/tarball/HEAD' a échoué
```

```
https://cran.rstudio.com/bin/windows/Rtools/
Warning in install.packages :
  impossible d'accéder à l'index de l'entrepôt
https://cran.rstudio.com/src/contrib:
  impossible d'ouvrir l'URL 'https://cran.rstudio.com/src/contrib/PACKAGES'
Warning in install.packages :
  le package 'remotes' n'est pas disponible for this version of R
```

Une version de ce package pour votre version de R est peut-être disponible ailleurs,
Voyez des idées à
<https://cran.r-project.org/doc/manuals/r-patched/R-admin.html#Installing-packages>
Warning in install.packages :
 impossible d'accéder à l'index de l'entrepôt
https://cran.rstudio.com/bin/windows/contrib/4.3:
 impossible d'ouvrir l'URL
'https://cran.rstudio.com/bin/windows/contrib/4.3/PACKAGES'

Exercice : chargement du package et de la page d'aide

```
# Chargement du package intro2r
library(intro2r)
# Affichage de la page d'aide
?fruits
```

Chargement des données fruits

Les fonctions utiles :

```
table(fruits$groupe)
barplot(table(fruits$groupe))
```

fruits

Exercice : extraction de données

Construisez la sous-table contenant la teneur en protéines, en glucides et en lipides des fruits secs.

```
new <- fruits[fruits$groupe == "secs", 5:7 ]
new
subset(fruits, groupe == "secs", select = c(Proteines, Glucides, Lipides))
fruits_extraction <- fruits[
  fruits$groupe == "secs",
  c("Proteines", "Lipides", "Glucides")]
fruits_extraction
```

```
> data("fruits",package="intro2")
Error in find.package(package, lib.loc, verbose = verbose) :
  aucun package nommé 'intro2' n'est trouvé
```

Explication : c'est "intro2r"

```
table.fruits <- subset(fruits$Proteines,fruits$Glucides,fruits$Lipides)
Erreur dans subset.default(fruits$Proteines, fruits$Glucides, fruits$Lipides)
:
  'subset' doit être une valeur logique
```

Cf solution ci-dessus

```
fruits$sous_table<-fruits[fruits$groupe=="secs",5:7]
```

Les fonctions utiles :

```
table(fruits$groupe)
barplot(table(fruits$groupe))
```

```
Error in `<-data.frame`(`*tmp*`, sous_table, value = list(Proteines =  
c(2.88, :
```

```
replacement has 5 rows, data has 51
```

Explication : pb de dimensions

```
fruitsecs <- fruits[fruits$groupe=="secs",fruits$Proteines,fruits$Glucides,fruits$Lipides]
```

```
Error in `[.data.frame`(fruits, fruits$groupe == "secs", fruits$Proteines, :  
unused argument (fruits$Lipides)
```

Exemple plus complexe avec subset

```
subset(fruits,  
       (groupe == "secs" &  
        Proteines >= 2) |  
       (groupe == "exotique" &  
        Sucres <= 50),  
       select = c(Proteines, Lipides, Glucides))
```

avec les crochets

```
fruits[(fruits$groupe == "secs" &  
        fruits$Proteines >= 2) |  
        (fruits$groupe == "exotique" &  
        fruits$Sucres <= 50),  
        c("Proteines", "Lipides", "Glucides")]
```

Explication : les fruits secs dont la teneur en protéines est sup à 2 ou bien les fruits exotiques dont la teneur en sucres est inférieure à 50

```
> subset(fruits,(groupe=="secs"&Proteine>=2),select=c(Lipides,Glucides))
```

```
Error in eval(e, x, parent.frame()) : object 'Proteine' not found
```

Les fonctions utiles :

```
table(fruits$groupe)
```

```
barplot(table(fruits$groupe))
```

```
fruits_secs_v4 <- subset(fruits, groupe=="secs" & Proteines>2) |
  (fruits, groupe=="exotique" & Sucres <= 50) ,
  select=c("Proteines", "Lipides", "Glucides"))
```

Error: unexpected ',' in:

```
"fruits_secs_v4 <- subset(fruits, groupe=="secs" & Proteines>2) |
  (fruits,"
```

Explication : problème de syntaxe

```
> subset(fruits,(groupe=="secs"&Proteine>=2),select=c(Lipides,Glucides))
```

Error in eval(e, x, parent.frame()) : object 'Proteine' not found

Explication : typo dans le nom de la colonne

RR

```
extraction3<-fruits[groupe=="secs" & Proteines<=2],select=c("Proteines", "Glucides",
"Lipides")]
```

Erreur : ')' inattendu(e) dans "extraction3<-fruits[groupe=="secs" & Proteines<=2]"

Explication :: mélange des deux syntaxes [] ou subset

```
extraction3 <- fruits[ fruits$groupe=="secs" & fruits$Proteines<=2, c("Proteines", "Glucides",
"Lipides")]
```

```
extraction3_bis <- subset(fruits, groupe=="secs" & Proteines<=2, select = c("Proteines",
"Glucides", "Lipides"))
```

```
zincetprot <- subset(fruits,(Zinc=>0.05)|(Proteines>1))
```

Error: '=>' is disabled; set '_R_USE_PIPEBIND_' envvar to a true value to enable it (<input>:1:34)

Explication : => plutôt que >= *Merci*

```
> db<-fruits[, c(Glucides, Lipides, groupe)]
```

Error: object 'Glucides' not found

Explication : il manque les guillemets

Les fonctions utiles :

```
table(fruits$groupe)
```

```
barplot(table(fruits$groupe))
```

```
Fruit_extraction <- subset(fruits, (groupe == "secs" & Proteines = 2) |  
(groupe == "exotique" & Sucres <= 50), c("Proteines", "Lipides",  
"Glucides"))  
Erreur dans groupe == "secs" & Proteines = 2 :  
  impossible de trouver la fonction "&<-"  
## = plutôt que == dans une opération logique
```

Exercice : estimation ponctuelle

Calculez

- la médiane de la teneur en sucres
- la moyenne de la teneur en eau
- l'écart-type de la teneur en eau

Appliquez la fonction `summary` aux données `fruits`.

- Calculez la corrélation de Pearson entre la teneur en eau et la teneur en sucres,
- Calculez la corrélation de Spearman

```
#la médiane de la teneur en sucres  
median(fruits$Eau)  
#la moyenne de la teneur en eau  
mean(fruits$Sucres)  
#l'écart-type de la teneur en eau  
sd(fruits$Eau)  
#Appliquez la fonction summary aux données fruits.  
summary(fruits)  
#Calculez la corrélation  
cor(fruits$Eau, fruits$Sucres)
```

Les fonctions utiles :

```
table(fruits$groupe)  
barplot(table(fruits$groupe))
```

```
> cor(fruits$Eau, fruits$sucre)
Error in cor(fruits$Eau, fruits$sucre) :
  supply both 'x' and 'y' or a matrix-like 'x'
## Explication : fruits$Sucre
```

Erreur de Spearman

```
cor.test(fruits$Eau, fruits$Sucre,
         alternative = c("two.sided", "less", "greater"),
         method = c("spearman"),
         exact = NULL, conf.level = 0.95, continuity = FALSE)
```

Warning message:

```
In cor.test.default(fruits$Eau, fruits$Sucre, alternative = c("two.sided", :
  Cannot compute exact p-value with ties
## Explication : il y a des valeurs ex-aequo dans la table
```

```
-----
Est-ce ok? (pas de message d'erreur)
?cor
cor(fruits$Eau, fruits$Glucides, use = "everything", method=c("pearson"))
cor(fruits$Eau, fruits$Glucides, use = "everything", method=c("spearman"))
-----
```

```
hist(fruits$Eau)
```

Mercredi après-midi : ggplot2

```
# Chargement du package ggplot2
library(ggplot2)
# Chargement des données
```

Les fonctions utiles :

```
table(fruits$groupe)
barplot(table(fruits$groupe))
```

```
data("fruits", package = "intro2r")
```

```
## Nuage de points
```

```
ggplot(  
  data = fruits,  
  mapping = aes(x = Sucres,  
                 y = Energie,  
                 color = groupe)) +  
  geom_point() +  
  geom_smooth(method = "lm")
```

Exercice : code à trous, diagramme en bâtons

À quoi servent les deux premières lignes ? Comment corriger la commande suivante pour obtenir le graphe à droite ?

idris : créer une nouvelle colonne dans la table dans laquelle sera indiqué TRUE ou FALSE selon que la teneur en sucre soit >10 ou non

```
fruits$SucresSup10 <- fruits$Sucres > 10  
ggplot(***, aes(***, fill = SucresSup10)) + geom_***()
```

```
ggplot(data=fruits, aes(x=SucresSup10, y=count, fill = SucresSup10)) + geom_bar()
```

Error in `geom_bar()`:

! Problem while computing aesthetics.

i Error occurred in the 1st layer.

Caused by error:

! object 'count' not found

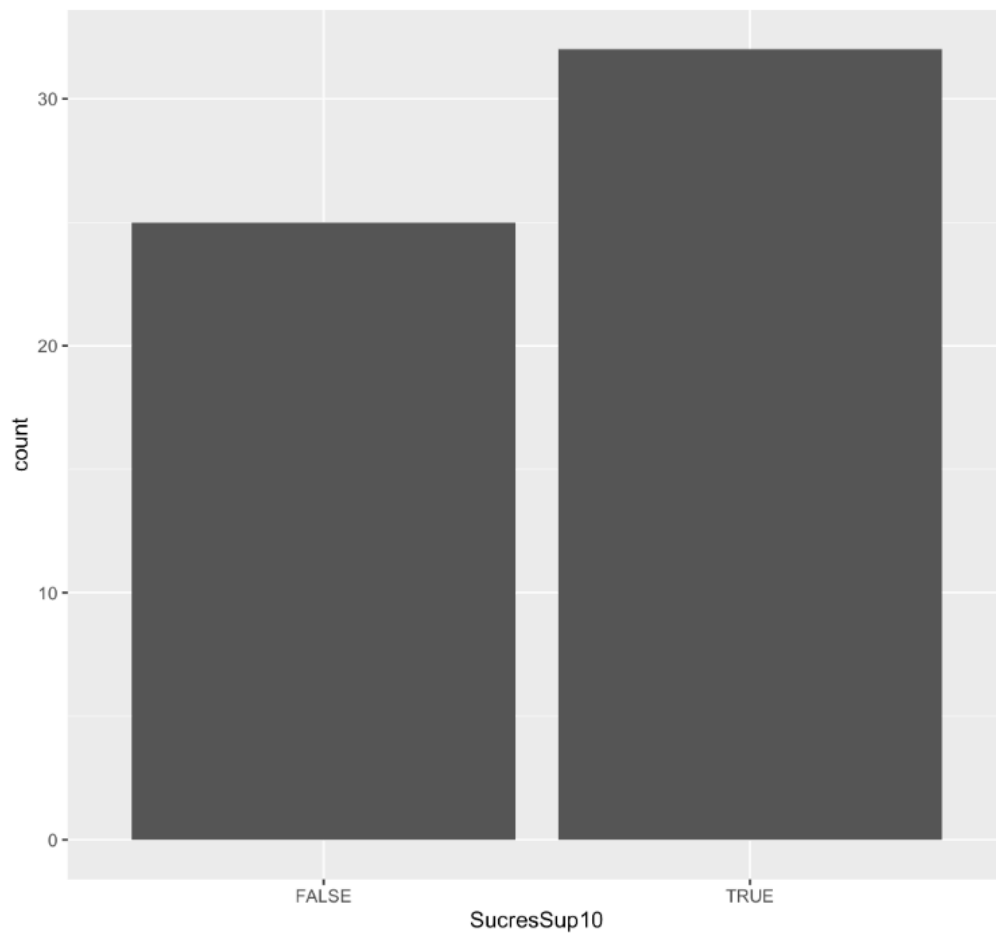
Run `rlang::last_trace()` to see where the error occurred.

Explication : la colonne "count" n'existe pas

Les fonctions utiles :

```
table(fruits$groupe)  
barplot(table(fruits$groupe))
```

```
ggplot(data = fruits, mapping = aes(x = SucresSup10)) + geom_bar()
```



```
ggplot(data=fruits, aes(fruits$SucresSup10, fill = SucresSup10)) + geom_bar()
```

Les fonctions utiles :

```
table(fruits$groupe)  
barplot(table(fruits$groupe))
```



```
> fruits$SucreSup10 <- fruits$Sucre > 10
Error: object 'fruits' not found
> ggplot(fruits, aes(SucreSup10, fill=SucreSup10)) + geom_bar()
Error in ggplot(fruits, aes(SucreSup10, fill = SucreSup10)) :
  could not find function "ggplot"
# Explication : il faut d'abord charger les données fruits (cf. plus haut)
```

Exemple avec geom_col

```
vitamine <- data.frame(
  groupe = c("compote", "crus", "exotique", "secs"),
  VitamineC = c(14.2, 12.2, 32.1, 0.3))

ggplot(data = vitamine, aes(x = groupe, y = VitamineC)) +
  geom_col()
```

Exercice geom_col :

Comment modifier le code précédent pour faire un diagramme en bâtons montrant le nombre de fruits ayant une teneur en Eau supérieure à 80 g / 100 g par groupe ?

(PS : il y a plus d'une solution possible)

```
fruits$EauSup80 <- fruits$Eau > 80
ggplot(fruits, mapping=aes(x=EauSup80)) + geom_bar()
```

Les fonctions utiles :

```
table(fruits$groupe)
barplot(table(fruits$groupe))
```

```
eau <- data.frame(
  groupe = c("compote", "crus", "exotique", "secs"),
  Eau = c(2, 29, 7, 0))
```

```
ggplot(data = eau , aes(x = groupe, y = Eau)) +
  geom_col()
```

```
fruits$EauSup80<-fruits$Eau>80
> ggplot(fruits$Eau, aes(x = groupe, y = EauSup80)) +geom_bar()
Error in `fortify()`:
! `data` must be a <data.frame>, or an object coercible by `fortify()`, or a
valid <data.frame>-like object
  coercible by `as.data.frame()`.
Caused by error in `.prevalidate_data_frame_like_object()`:
! `dim(data)` must return an <integer> of length 2.
Run `rlang::last_trace()` to see where the error occurred.
# Explication: le premier argument de ggplot doit être la data-frame
```

Charlotte:

```
ggplot(data=fruits, aes(x=Eau>=80, fill=groupe)) + geom_bar(position="dodge",
color="white") + ylim(0,40) +labs (x="Teneur en Eau", y="Fréquence")
```

Anissa :

```
table(fruits$Eausup80, fruits$groupe)
```

Les fonctions utiles :

```
table(fruits$groupe)
barplot(table(fruits$groupe))
```

```
ggplot(fruits, aes(x = groupe, y = EauSup80)) +geom_bar()
Error in `geom_bar()` :
! Problem while computing stat.
! Error occurred in the 1st layer.
Caused by error in `setup_params()` :
! `stat_count()` must only have an x or y aesthetic.
Run `rlang::last_trace()` to see where the error occurred.
```

>

Solution plus complexe

```
count_eau <- tapply(
  fruits$Eau > 80,
  fruits$groupe,
  sum)
```

```
eau <- data.frame(
  groupe = names(count_eau),
  n = count_eau
)
```

```
ggplot(eau, aes(groupe, n)) + geom_col()
```

Les fonctions utiles :

```
table(fruits$groupe)
barplot(table(fruits$groupe))
```

Mahdi

```
fruits$fruits80 <- (fruits$Eau>80)
```

```
ggplot(data=fruits,mapping=aes(x=fruits80, y=groupe,fill=fruits80))+geom_bar())
```

Error in `ggplot()`:

! `mapping` must be created with `aes()`.

✖ You've supplied NULL.

Run `rlang::last_trace()` to see where the error occurred.

souci de parenthèses

Anissa :

```
count_eau <- tapply (fruits$Eau>80, fruits$groupe, sum)
```

```
eau<-data.frame(groupe=(count_eau), n=count_eau)
```

```
ggplot(eau, aes(groupe, n)) + geom_col()
```

je n'ai pas les noms des groupes...

```
fruits$EauSup80 <- fruits$Eau > 80
```

```
library(dplyr)
```

```
Nombre_fruits <- fruits %>%
```

```
  group_by(groupe) %>%
```

```
  summarize(Nombre = sum(EauSup80))
```

```
ggplot(Nombre_fruits, aes(x=groupe, y=Nombre)) +
```

Les fonctions utiles :

```
table(fruits$groupe)
```

```
barplot(table(fruits$groupe))
```

```
geom_bar(stat="identity") +
labs(x = "Groupes", y = "Teneur en Eau > 80g/100g", title = "Teneur en Eau par Groupe") +
theme_minimal()
```

```
fruits_filtered <- subset(fruits, Eau > 80)
# Create the plot
#have it be by group
ggplot(data = fruits_filtered, aes(x = groupe)) +
  geom_bar()
```

```
eaufruits <- table(fruits$groupe, fruits$fruits80)
eaufruits
eaufruits <- as.data.frame(eaufruits)
ggplot(data=eaufruits, mapping=aes(x=Var2, y=Freq, fill=Var2))+geom_bar()
Error in `ggplot()` :
! `mapping` must be created with `aes()`.
✖ You've supplied NULL.
Run `rlang::last_trace()` to see where the error occurred.
```

```
eaufruits <- table(fruits$groupe, fruits$fruits80)
eaufruits <- subset(
  as.data.frame(eaufruits), Var2 == TRUE)
ggplot(
  data=eaufruits,
  mapping=aes(x=Var1, y=Freq, fill=Var1))+
  geom_col()
```

Les fonctions utiles :

```
table(fruits$groupe)
barplot(table(fruits$groupe))
```

Exercice : histogramme de la teneur en Vitamine C

```
# exemple sur la teneur en sucres
ggplot(fruits, aes(Sucres)) +
  geom_histogram(breaks = seq(0, 75, 5),
    fill = "steelblue",
    color = "white")
```

Laure

```
ggplot(fruits, aes(VitamineC)) +
  geom_histogram(breaks = seq(0, 85, 5),
    fill = "red",
    color = "white")
```

```
ggplot(fruits, aes(VitamineC)) +
  geom_histogram(breaks = seq(0, 75, 5),
    fill = "orange",
    color = "red")
```

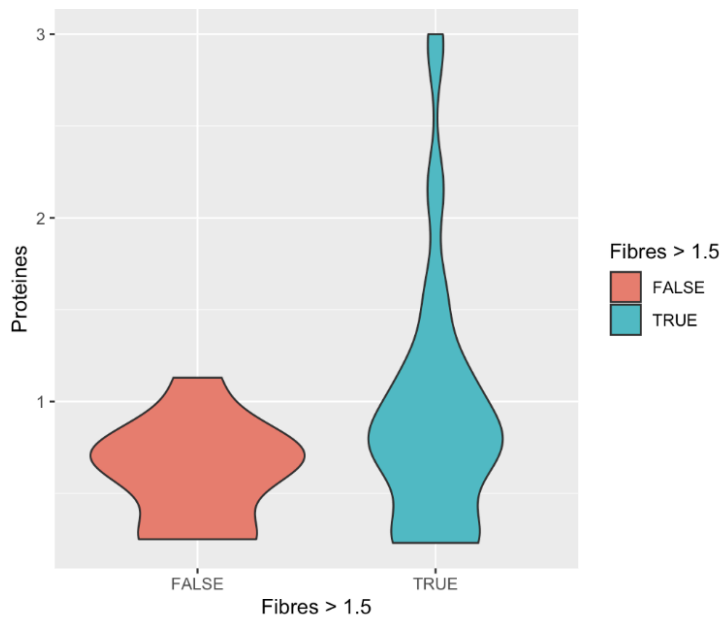
Exercice à trous : diagramme en violons

Complétez le code suivant pour obtenir la figure de droite :

```
ggplot(fruits,
  aes(x = Fibres > 1.5,
    y = Proteines,
    fill = ***) +
  geom_***())
```

Les fonctions utiles :

```
table(fruits$groupe)
barplot(table(fruits$groupe))
```



```
ggplot(fruits,
  aes(x = Fibres > 1.5,
    y = Proteines,
    fill = Fibres > 1.5)) +
  geom_violin()
```

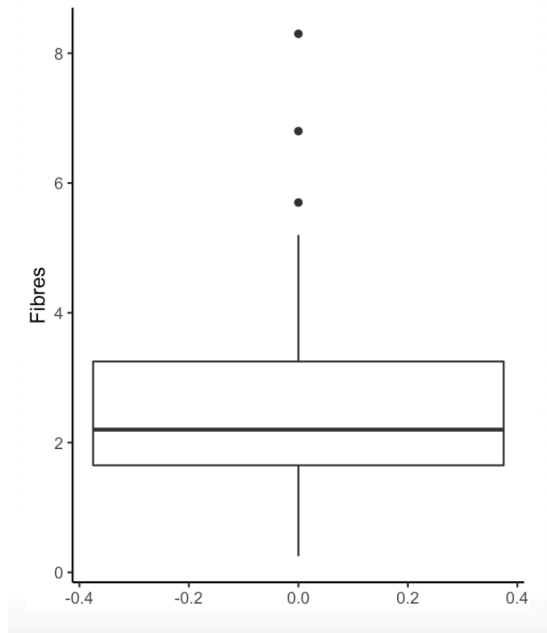
Exercice à trous : theme

1. Consultez la page d'aide de la fonction `theme_bw` avec la commande `?theme_bw`
2. Choisissez le thème permettant de réaliser le graphe à droite en complétant la commande suivante.

Les fonctions utiles :

```
table(fruits$groupe)
barplot(table(fruits$groupe))
```

```
ggplot(fruits, aes(y = Fibres)) +
  geom_boxplot() +
  theme_***()
```



```
ggplot (fruits, aes (y=Fibres))+ geom_boxplot() + theme_classic ()
```

```
ggplot(fruits, aes(y = Fibres)) +
  geom_boxplot() +
  theme_test()
```

```
ggplot(fruits, aes (y=Fibres))+
```

Les fonctions utiles :

```
table(fruits$groupe)
barplot(table(fruits$groupe))
```



```

geom_hline(
  yintercept = 4,
  color = "steelblue",
  linetype = "dashed", size = 1) +
geom_boxplot(
  fill = NA,
  size = 1,
  color = "steelblue") +
theme_classic() +
theme(
  axis.title = element_text(
    size = 20,
    face = "bold",
    color = "darkblue"),
  axis.ticks = element_line(
    color = "grey"),
  axis.ticks.length = unit(.25, "cm"),
  panel.background = element_rect(
    fill = "#0355994E")
)

```

Mahdi

```

ggplot(fruits,aes(y=Fibres))+geom_boxplot()+theme_classic()+theme(axis.title=element_text(
size=20,face="bold",color="darkblue")

```

```

"ggplot(fruits,aes(y=Fibres))+geom_boxplot()+theme_classic()+theme(axis.title
=element_text(size=20,face="bold",color="darkblue")
ggplot"

```

```

ggplot(
  data = fruits,

```

Les fonctions utiles :

```

table(fruits$groupe)
barplot(table(fruits$groupe))

```

```

mapping = aes(
  x = groupe,
  fill = groupe)) +
geom_bar() +
scale_fill_manual(
  values = c("orange",
    "yellow",
    "limegreen",
    "brown"))

# Personnalisation des couleurs
ggplot(fruits, aes(Phosphore,
  Calcium)) + geom_point(color = "white") +
scale_x_log10() +
scale_y_log10() +
labs(x = "log10(Phosphore)",
  y = "log10(Calcium)") + theme_dark()

```

Exercice à trous : personnalisation

Complétez la commande suivante pour obtenir le graphe ci-contre.

```

ggplot(fruits,
  aes(Phosphore,
    Calcium)) +
geom_point(*** = "white") +
scale_***() +
scale_***() +
labs(x = "log10(Phosphore)",
  y = "log10(Calcium)") +
theme_dark()

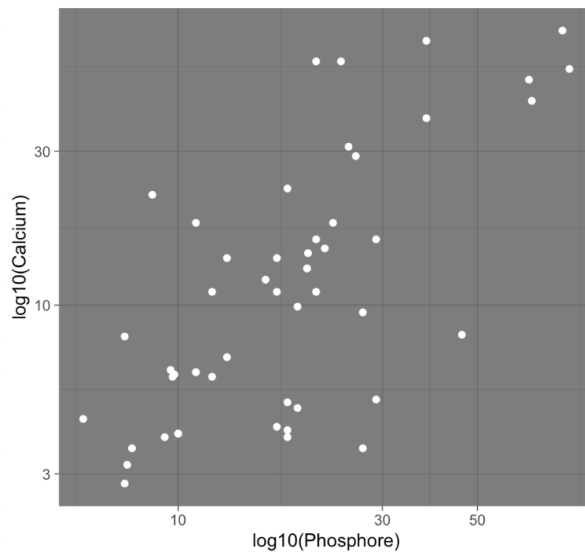
```

Les fonctions utiles :

```

table(fruits$groupe)
barplot(table(fruits$groupe))

```



```
ggplot(fruits, aes(Phosphore,
                    Calcium)) +
  geom_point(color = "white") +
  scale_x_log10() +
  scale_y_log10() +
  labs(x = "log10(Phosphore)",
       y = "log10(Calcium)") + theme_dark()
```

Anissa

```
ggplot(fruits,
       aes(Phosphore,
           Calcium)) +
  geom_point(color = "white") +
  scale_x_log10() +
  scale_y_log10() +
```

Les fonctions utiles :

```
table(fruits$groupe)
barplot(table(fruits$groupe))
```

```
labs(x = "log10(Phosphore)",  
      y = "log10(Calcium)") +  
theme_dark()
```

```
g <- ggplot(fruits, aes(groupe)) +  
  geom_bar()  
g  
ggsave(  
  filename = "graphe_batons.png",  
  plot = g)
```

Jeudi Matin : visualisation avec ggplot2

Démonstration sur les données fruits

```
# Chargement des librairies  
library(readxl)  
library(ggplot2)  
  
# Chargement des données  
fruits <- read_excel("fruits.xlsx")  
  
# Représentation graphique  
g <- ggplot(  
  data = fruits,  
  mapping = aes(  
    x = groupe,  
    y = Energie,
```

Les fonctions utiles :

```
table(fruits$groupe)  
barplot(table(fruits$groupe))
```

```

    fill = groupe)
) +
geom_boxplot(
  size = 0.9,
  alpha = 0.5) +
geom_point(aes(color = groupe), size = 2) +
scale_y_log10() +
labs(y = "Energie (log-scale)") +
theme_bw()

plot(g)

```

```

# Sauvegarde du graphe
ggsave(filename = "boxplot.png", plot = g)

```

```

> library(readxl)
> fruits <- read_excel("fruits.xls")
Error: `path` does not exist: 'fruits.xls'

```

Bastien

```

ggplot(data = fruits, mapping = aes
  (x = groupe, y = Energie, fill = groupe)
)

+ geom_boxplot(size = 0.9, alpha = 0.5) +
  geom_point(aes(color = groupe)) +
  scale_y_log10()+
  labs ( y ="Energie (log-scale)"

```

Les fonctions utiles :

```

table(fruits$groupe)
barplot(table(fruits$groupe))

```

```
) +  
theme_bw()
```

```
Error in `+.gg`:  
! Cannot use `+` with a single argument.  
i Did you accidentally put `+` on a new line?  
Run `rlang::last_trace()` to see where the error occurred.  
# Explication : attention à bien mettre le "+" à la fin de la ligne
```

```
> g<-ggplot  
> plot(g)  
Error in `fortify()``:  
! `data` must be a <data.frame>, or an object coercible by `fortify()``,  
or a valid <data.frame>-like object coercible by `as.data.frame()``.  
Caused by error in `.prevalidate_data_frame_like_object()``:  
! `dim(data)` must return an <integer> of length 2.  
Run `rlang::last_trace()` to see where the error occurred.
```

Jeudi après-midi : ggplot (fin) et tests

Joël

#Chargement des librairies

```
library(readxl)
```

```
library(ggplot2)
```

#Chargement des données

```
nutrimenu <- read_excel("nutrimenu.xlsx")
```

#Réalisation des graphiques

Les fonctions utiles :

```
table(fruits$groupe)
```

```
barplot(table(fruits$groupe))
```

#1-Histogrammes

```
graph1 <- ggplot(  
  data = nutrimenu,  
  mapping = aes(  
    x=Nutriscore,  
    fill=Type))+  
  geom_bar()+  
  theme_bw()  
graph1
```

#teneur en energie

```
summary(nutrimenu$Energie)  
graph2 <- ggplot(  
  data = nutrimenu,  
  mapping = aes(Energie)) +  
  geom_histogram(  
    breaks = seq(0, 400, 50),  
    fill = "pink",  
    color = "black")  
graph2
```

#Type de recette > 50 kCal

```
EntreeSup50 <- subset(  
  nutrimenu,  
  (Type == "Entree" & Energie >= 50),  
  select = c(Proteines, Lipides,  
    Glucides,Sucre,Fibres,  
    Energie))  
PlatSup50 <- subset(  
  nutrimenu,  
  (Type == "Plat" & Energie >= 50),  
  select = c(Proteines, Lipides,
```

Les fonctions utiles :

```
table(fruits$groupe)  
barplot(table(fruits$groupe))
```

```

      Glucides, Sucre, Fibres,
      Energie))
DessertSup50 <- subset(
  nutrimenu,
  (Type == "Dessert" & Energie >= 50),
  select = c(Proteines, Lipides,
    Glucides, Sucre, Fibres,
    Energie))

```

```

graph3_entrees <- ggplot(
  data = EntreeSup50,
  mapping = aes(x = Energie))+
  geom_histogram(
    breaks = seq(0, 400, 50))+
  theme_bw()

```

```

graph3_plats <- ggplot(
  data = PlatSup50,
  mapping = aes(x = Energie))+
  geom_histogram(
    breaks = seq(0, 400, 50))+
  theme_bw()

```

```

graph3_desserts <- ggplot(
  data = DessertSup50,
  mapping = aes(x = Energie))+
  geom_histogram(
    breaks = seq(0, 400, 50))+
  theme_bw()

```

```

graph3_entrees
graph3_plats
graph3_desserts

```

Les fonctions utiles :

```

table(fruits$groupe)
barplot(table(fruits$groupe))

```



```
#Graphe en point Energie_Glucides
```

```
graph4 <- ggplot(  
  data = nutrimenu,  
  mapping = aes(  
    x = Energie,  
    y = Glucides)) +  
  geom_point(color = "blue") +  
  scale_x_log10() +  
  scale_y_log10() +  
  labs(x = "(Energie)",  
       y = "(Glucides)") +  
  theme_dark()  
graph4
```

```
#Boxplot Energie
```

```
r_Energie <- ggplot(  
  data = nutrimenu,  
  mapping = aes(  
    x = Type,  
    y = Energie,  
    fill = Type))+  
  geom_boxplot(size=1, alpha=0.5)+  
  theme_bw()+  
  scale_y_log10()+  
  labs(y = "Energie (log-scale)") +  
  geom_point(aes(color=Type), size=1)  
r_Energie
```

```
#Teneur en sucre>5
```

```
graph5 <- ggplot(  
  data = nutrimenu,  
  mapping = aes(  
    x = Type,  
    y = Teneur_sucres)) +  
  geom_bar(aes(fill = Type)) +  
  theme_minimal() +  
  labs(x = "Type", y = "Teneur_sucres") +  
  scale_y_log10() +  
  scale_x_discrete(limits = rev(levels(nutrimenu$Type))) +  
  theme(axis.text.x = element_text(angle = 45, hjust = 1))  
graph5
```

Les fonctions utiles :

```
table(fruits$groupe)  
barplot(table(fruits$groupe))
```

```

data = nutrimenu,
mapping = aes(
  x = Sucre > 5,
  y = Glucides,
  fill = Sucre > 5)) +
geom_violin()
graph5

```

```

#Sauvegarde des graphiques
ggsave(filename = "distrub.png", plot = graph1)
ggsave(filename = "Energie.png", plot = graph2)
ggsave(filename = "Energ-Glu.png", plot = graph4)
ggsave(filename = "boxplotEnerg.png", plot = r_Energie)

```

```

ggplot(nutrimenu,
  aes(Type, Energie, color = Nutriscore)) +
stat_summary(
  position = position_dodge(width = 0.3)
) +
geom_point(
  position = position_dodge(width = 0.3)) +
theme_bw()

```

Alexisse

```

ggplot (data = nutrimenu, mapping = aes(x = Nutriscore, y = Glucides, fill= Type)) +
geom_boxplot(size = 0.7, alpha = 0.4) + labs(title = "Teneur en Glucides en fonction du
Nutriscore") + theme_classic () + scale_fill_manual (values = c("orange","green","blue"))

```

```

ggplot (data = nutrimenu, mapping = aes(x = Nutriscore, y = Glucides, fill= Type)) +
geom_boxplot(size = 0.7, alpha = 0.4) + labs(title = "Teneur en Glucides en fonction du

```

Les fonctions utiles :

```

table(fruits$groupe)
barplot(table(fruits$groupe))

```

```
Nutriscore") + theme_classic () + scale_fill_manual (values = c("orange","green","blue")) +  
geom_signif (comparisons = list (c("A", "B")))
```

```
>  
ggplot (nutrimenu, aes (Glucides, fill=Type)) + geom_histogram (breaks=seq (0, 50, 10),  
fill="pink") )  
Error: unexpected ')' in  
"ggplot (nutrimenu, aes (Glucides, fill=Type)) + geom_histogram (breaks=seq (0, 50, 10)  
, fill="pink") ) "  
# Explication : la dernière parenthèse fermante est superflue
```

Anissa

```
ggplot(nutrimenu, aes(x=Nutriscore, y=Glucides, fill=Type)) +  
  geom_boxplot(alpha=0.2, color="White") +  
  geom_point(alpha= 0.5, size =2, fill= "Blue") +  
  theme_classic() +  
  labs(y= "Teneur en glucides")
```

```
#installation package: pak pour installer package  
install.packages("pak")  
library(pak)
```

```
pak::pak("ggppubr")
```

```
> pak::pak("ggppubr")  
✓ Updated metadata database: 5.29 MB in 15 files.  
✓ Updating metadata database ... done  
Error:  
! error in pak subprocess  
Caused by error:
```

Les fonctions utiles :

```
table(fruits$groupe)  
barplot(table(fruits$groupe))
```

```
! Could not solve package dependencies:  
* ggppubr: Can't find package called ggppubr.  
Type .Last.error to see the more details.  
# Explication : c'est ggpubr
```

Maite

Do you want to install from sources the package which needs compilation? (Yes/no/cancel)
install.packages("pak")
Error in install.packages : Unrecognized response "install.packages("pak")"

```
## installation avec pak  
install.packages("pak")  
## Installation du package ggpubr avec pak  
pak::pak("ggpubr")
```

```
library(ggpubr)
```

```
> ggplot(nutrimenu,aes(Nutriscore,Glucides,fill=Type))+  
+   geom_boxplot(aes(size=0.7,alpha=0.4)+geom_signif(comparisons =  
list(c("A","B"))  
+ ggplot(nutrimenu,aes(Nutriscore,Glucides,fill=Type))+  
Error: unexpected symbol in:  
"   geom_boxplot(aes(size=0.7,alpha=0.4)+geom_signif(comparisons =  
list(c("A","B"))  
ggplot"
```

Explication : corriger les parenthèses et faire attention au prompt

```
ggplot(nutrimenu,aes(Nutriscore,Glucides,fill=Type)) +  
  geom_boxplot(size=0.7,alpha=0.4)+geom_signif(comparisons =  
list(c("A","B")))
```

Les fonctions utiles :

```
table(fruits$groupe)  
barplot(table(fruits$groupe))
```

Mahdi

```
ggplot(data=nutrimenu, mapping=aes(x=Nutriscore,y=Energie,fill=Type))+  
  geom_boxplot(alpha=0.5) +  
  theme_bw() +  
  labs(x="Nutriscore",y="Energie(kcal)", title="Quantité d'énergie par plat en fonction du  
Nutriscore et du type de plat")+  
  scale_fill_manual(values=c("chocolate","darkgreen","red"))+  
  
geom_signif(comparisons=list(c("A","B"),c("A","C"),c("B","C"),c("Dessert","Entree")),step_incr  
ease = 0.1)  
Caused by error in `if (scales$x$map(comp[1]) == data$group[1] | manual)  
...`:  
! missing value where TRUE/FALSE needed
```

Pour les designs plus complexes :

<https://www.datanovia.com/en/blog/how-to-add-p-values-onto-a-grouped-ggplot-using-the-ggpubr-r-package/>

ANOVA

Chargement des librairies

```
library(ggplot2)
```

Chargement des données

```
fruits <- read.csv("fruits.csv", skip = 2)
```

Représentation graphique

```
ggplot(
```

Les fonctions utiles :

```
table(fruits$groupe)
```

```
barplot(table(fruits$groupe))
```

```
data = fruits,  
mapping = aes(  
  x = groupe,  
  y = VitamineC,  
  color = groupe)) +  
geom_boxplot(fill = NA, size = 0.8) +  
theme_minimal()
```

```
# ANOVA  
summary(aov(VitamineC ~ groupe, data = fruits))  
# ANOVA non paramétrique  
kruskal.test(VitamineC ~ groupe, data = fruits)
```

```
# Pivot  
library(tidyr)  
fruits %>%  
  pivot_longer(-(1:2)) %>%  
  ggplot(aes(  
    value, reorder(name, value, median))) +  
  geom_boxplot() +  
  theme_bw() +  
  scale_x_log10()
```

Recap

```
## Chargement des librairies  
library(dplyr)  
library(ggpubr)  
library(ggplot2)
```

Les fonctions utiles :

```
table(fruits$groupe)  
barplot(table(fruits$groupe))
```

```

## Chargement des données
fruits <- read.delim2("fruits.txt")

## Compter le fruits
table(fruits$groupe)
prop.table(table(fruits$groupe))

## Subset
cruexo_sub <- subset(
  fruits,
  (groupe == "exotique") |
    (groupe == "crus"),
  select = c(nom, groupe, VitamineC))

## Crochets carrés
cruexo_cc <- fruits[
  (fruits$groupe == "exotique") |
    (fruits$groupe == "crus"),
  c("nom", "groupe", "VitamineC")]

## Dplyr
cruexo_dplyr <- fruits %>%
  filter(
    (groupe == "exotique") |
      (groupe == "crus")) %>%
  select(nom, groupe, VitamineC)

## Graphe
ggplot(
  cruexo_sub, aes(
    groupe,
    VitamineC,

```

Les fonctions utiles :

```

table(fruits$groupe)
barplot(table(fruits$groupe))

```

```

    fill = groupe)) +
  geom_boxplot() +
  geom_signif(
    comparisons = list(
      c("crus", "exotique")),
    test = "t.test") +
  scale_y_log10() +
  theme_classic()

```

Test de Student

```

t.test(
  cruexo_sub$VitamineC ~
  cruexo_sub$groupe)

```

```

> ggplot(cruexo, aes(groupe, VitamineC, fill = groupe)) + geom_col()
> ggplot(cruexo, aes(groupe, VitamineC, fill = groupe)) + geom_violin()
> library (ggpubr)
> ggplot(cruexo, aes(groupe, VitamineC, fill = groupe)) + geom_boxplot() + scale_y_log10() +
+   theme_classic() + geom_signif(comparisons = list(c("cru", "exotique")))
Warning message:
Computation failed in `stat_signif()`.
Caused by error in `if (scales$x$map(comp[1]) == data$group[1] | manual) ...`:
! missing value where TRUE/FALSE needed
# Explication : c'est "crus" (avec un s)

```

>

Les fonctions utiles :

```

table(fruits$groupe)
barplot(table(fruits$groupe))

```



```
cruexo_cc<-fruits[(fruits$groupe=="exotique"|(fruits$groupe=="crus"),  
                  c("nom","groupe","VitamineC")]
```

Explication : il manque une parenthèse fermante après "exotique"

```
> cruexo_cc <- fruits [(fruits$groupe=="exotique") |  
  (fruits$groupe=="exotique"), c("nom", "groupe", "vitamineC")]  
Erreur dans `[.data.frame`(fruits, (fruits$groupe == "exotique") |  
  (fruits$groupe ==  :  
  colonnes non définies sélectionnées
```

Explication : c'est une "V" (majuscule)

Les fonctions utiles :

```
table(fruits$groupe)  
barplot(table(fruits$groupe))
```