

Outils de debugging/débogage

Présentation des outils de débogage avec R et RStudio

Sommaire

- 1 Le print
- 2 Les breakpoints
- 3 La fonction browser
- 4 Les fonctions debug() / debugonce()
- 5 Exercice 1
- 6 L'option recover
- 7 Astuce dans les chaine de traitements
- 8 Le package boomer
- 9 Exercice 2
- 10 Les logs

0.1 Sommaire

- Pourquoi faire du debugging/débogage ?
- Le print
- Le breakpoint / point d'arrêt
- La fonction browser
- Les fonctions debug() / debugonce()
- L'option recover
- Astuce dans les chaine de traitements
- Le package boomer
- Les logs

0.2 Pourquoi faire du debugging/débogage ?

Le débogage, ou debugging en anglais, est une étape cruciale dans le développement de tout programme informatique.

Cette présentation vous présente quelques outils que vous avez déjà à disposition sous **votre RStudio sur Cerise** afin de vous aider à trouver par vous-même la source de vos bugs.

0.3 Exemple de code

Soit la fonction suivante :

```
1 exemple1 <- function(indicateur) {  
2  
3   if (indicateur == "Sepale") {  
4     resultat <- iris |>  
5       select(Species, starts_with("Sepal")) |>  
6       head(3)  
7   } else if (indicateur == "Pétale") {  
8     resultat <- iris |>  
9       select(Species, starts_with("Petal")) |>  
10      head(3)  
11   }  
12   return(resultat)  
13  
14 }
```



0.4 Des erreurs/bugs

```
1 ## Tout est ok
2 exemple1("Petale")
3
4 > exemple1("Petale")
5   Species Petal.Length Petal.Width
6 1  setosa         1.4         0.2
7 2  setosa         1.4         0.2
8 3  setosa         1.3         0.2
```

```
1 ## Un bug
2 exemple1("petale")
3
4 > exemple1("petale")
5 Erreur dans exemple1("petale") : objet 'resultat' introuvable
```

1 Le print



1.1 Les prints (1/3)

Il existe un outil très simple (universel car non spécifique à R), parfois décrié mais qui reste efficace : ajouter la fonction `print()` dans votre code R pour examiner le contenu d'un objet à un moment donné.

Par exemple :

```
1 exemple1 <- function(indicateur) {  
2  
3   print(indicateur) # <- on ajoute un print()  
4  
5   if (indicateur == "Sepale") {  
6     resultat <- iris |>  
7       select(Species, starts_with("Sepal")) |>  
8       head(3)  
9   } else if (indicateur == "Pétale") {  
10    resultat <- iris |>  
11      select(Species, starts_with("Petal")) |>  
12      head(3)  
13  }  
14  return(resultat)  
15  
16 }
```


1.2 Les prints (2/3)

Cela renvoie pour le cas du bug :

```
1 > exemple1("petale")  
2 [1] "petale"  
3 Erreur dans exemple1("petale") : objet 'resultat' introuvable
```

=> On récupère ainsi la valeur contenue dans l'objet indicateur "petale" ce qui nous permet de comprendre qu'on ne rentre dans aucune des 2 conditions de la structure conditionnelle de `exemple1()`.

=> L'objet `resultat` n'ayant pas été créé, R ne peut pas le trouver :).

1.3 Les prints (3/3)

👍 Avantages du print :

- Simplicité, universalité
- Pas d'interruption du code
- Compatible dans tous les environnements (RStudio, VSCode, batchs...)

👎 Inconvénients du print :

- Intrusif dans le code - Pas d'inspection de l'environnement à la volée
- On oublie souvent de les retirer du code une fois les bugs résolus.



2 Les breakpoints



2.1 Les breakpoints (1/6)

Avant de lancer vos traitements, il peut être utile de définir des breakpoints (“point d’arrêt”) sur une ligne de code. Lorsque R atteint cette ligne pendant l’exécution, il met le script en pause.

Pour cela, on peut cliquer dans la marge à gauche du numéro de ligne dans l’éditeur.

=> Cela permet d’avoir accès à l’environnement d’exécution, ce qui signifie que vous pouvez voir les valeurs de toutes vos variables, exécuter des commandes dans la console comme si vous étiez à cet endroit précis du code, et avancer pas à pas pour observer comment les variables changent.

```
3 exemple1 <- function(indicateur) {  
4  
5 if (indicateur == "Sepale") {  
6   resultat <- iris |>  
7     select(Species,starts_with("Sepal")) |>  
8     head(3)  
9 } else if (indicateur == "Pétale") {  
10  resultat <- iris |>  
11    select(Species,starts_with("Petal")) |>  
12    head(3)  
13 }  
14 return(resultat)  
15  
16 }
```

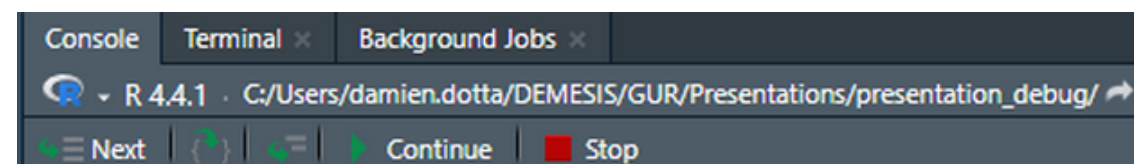
Les fonctions R qui contiennent un breakpoint ont un point rouge dans l’onglet Environnement de RStudio :



2.2 Les breakpoints (2/6)

En appuyant sur le bouton **source** de RStudio :

```
3 exemple1 <- function(indicateur) {  
4  
5   if (indicateur == "Sepale") {  
6     resultat <- iris |>  
7       select(Species, starts_with("Sepal")) |>  
8       head(3)  
9   } else if (indicateur == "Pétale") {  
10    resultat <- iris |>  
11      select(Species, starts_with("Petal")) |>  
12      head(3)  
13  }  
14  return(resultat)  
15  
16 }
```



La flèche verte et le texte en surbrillance permettent d'identifier le bout de code en cours d'exécution

Appuyer sur le bouton “Next” ou taper la lettre “n” dans la console de RStudio pour exécuter le script pas à pas.

2.3 Les breakpoints (3/6)

En appuyant sur le bouton **Next** de RStudio :

values	
indicateur	"Pétale"

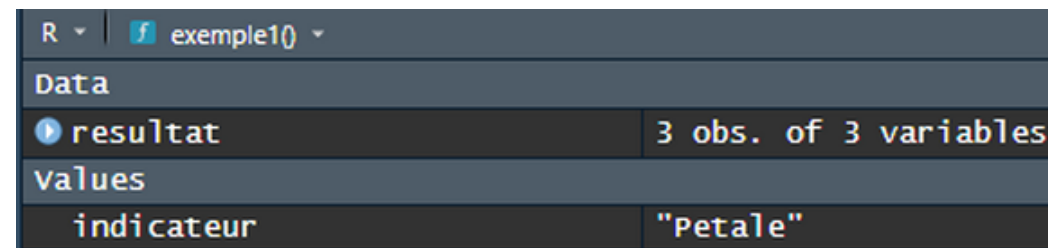
```
5  if (indicateur == "Sepale") {  
6    resultat <- iris |>  
7    select(species, starts_with("Sepal")) |>  
8    head(3)  
9  } else if (indicateur == "Pétale") {  
10   resultat <- iris |>  
11   select(species, starts_with("Petal")) |>  
12   head(3)  
13 }  
14 return(resultat)  
15  
16 }
```

Les objets se mettent à jour dans l'onglet Environnement de RStudio.

La flèche verte se positionne à l'emplacement du code qui sera exécuté à l'étape suivante.

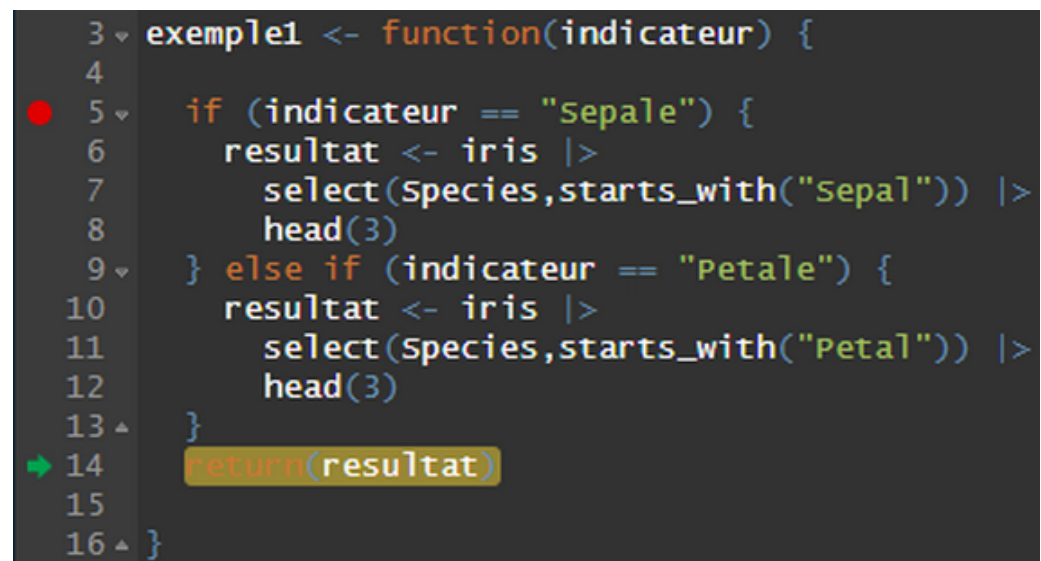
2.4 Les breakpoints (4/6)

En appuyant sur le bouton **Next** de RStudio :



The screenshot shows the RStudio Environment pane. At the top, it says 'R' and 'exemple1()'. Below that, under the 'Data' tab, there is an object named 'resultat' with the description '3 obs. of 3 variables'. Under the 'Values' tab, the variable 'indicateur' is shown with the value 'Pétale'.

Data	
▶ resultat	3 obs. of 3 variables
Values	
indicateur	"Pétale"



The screenshot shows the RStudio Source pane with a function definition. A red dot (breakpoint) is placed on line 5, which is the start of the 'if' statement. A green arrow points to line 14, which is the 'return(resultat)' statement.

```
3 exemple1 <- function(indicateur) {  
4  
5 if (indicateur == "sepale") {  
6   resultat <- iris |>  
7     select(species, starts_with("Sepal")) |>  
8     head(3)  
9 } else if (indicateur == "Pétale") {  
10  resultat <- iris |>  
11    select(species, starts_with("Petal")) |>  
12    head(3)  
13 }  
14 return(resultat)  
15  
16 }
```

Les objets se mettent à jour dans l'onglet Environnement de RStudio.

La flèche verte se positionne à l'emplacement du code qui sera exécuté à l'étape suivante.

2.5 Les breakpoints (5/6)

👍 Avantages des breakpoints :

- Très visuels
- Pas besoin de modifier votre code R
- Permet l'exécution pas à pas

👎 Inconvénients des breakpoints :

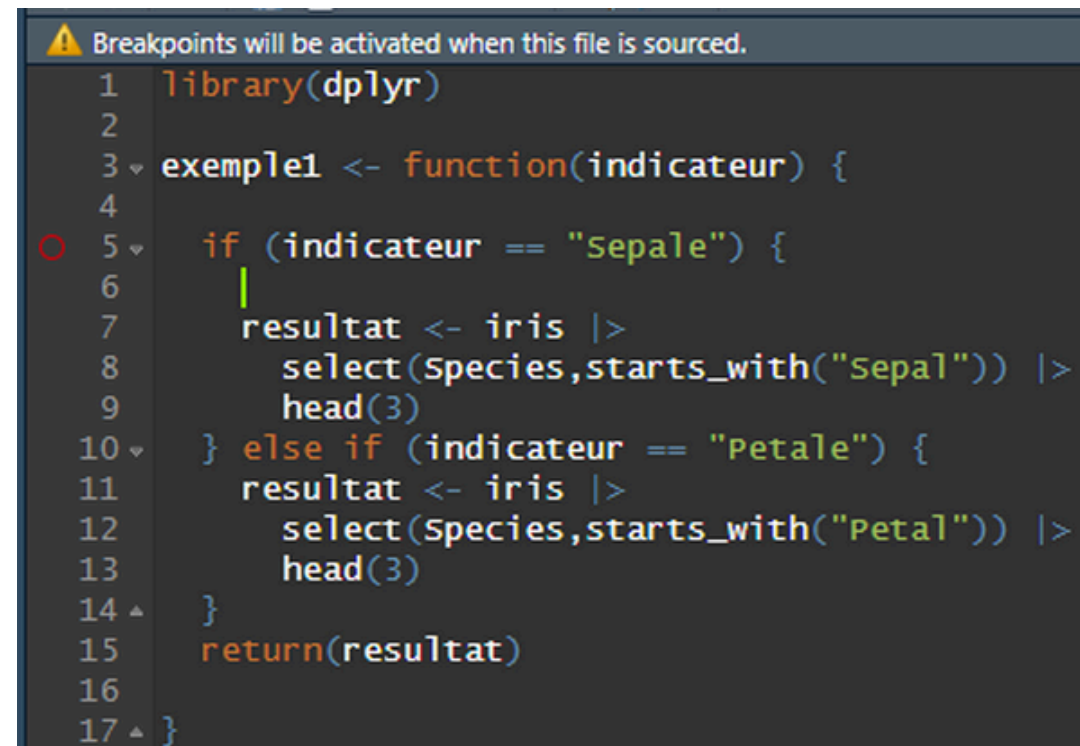
- Spécifique à RStudio et aux programmes R
- Pas versionnable avec Git
- Difficile à utiliser dans certains cas dynamiques (les shiny par ex.)



2.6 Les breakpoints (6/6)

⚠ Remarque

Si lorsque vous injectez un breakpoint dans votre programme R **celui-ci s'affiche en cercle rouge**, il faudra alors sauvegarder votre script et le sourcer pour retrouver le rond rouge.



```
1 library(dplyr)
2
3 exemple1 <- function(indicateur) {
4
5   if (indicateur == "Sepale") {
6     |
7     resultat <- iris |>
8       select(Species,starts_with("Sepal")) |>
9       head(3)
10  } else if (indicateur == "Petal") {
11    resultat <- iris |>
12      select(Species,starts_with("Petal")) |>
13      head(3)
14  }
15  return(resultat)
16
17 }
```

3 La fonction browser



3.1 La fonction browser (1/3)

La fonction browser() est un compromis entre la fonction print() et les breakpoints.

- Elle peut s'insérer à n'importe quel endroit du code comme print().
- Elle permet l'exécution pas à pas comme les breakpoints.



3.2 La fonction browser (2/3)

Par exemple :

```
1 exemple1 <- function(indicateur) {  
2  
3   browser() # <- on ajoute un browser()  
4  
5   if (indicateur == "Sepale") {  
6     resultat <- iris |>  
7       select(Species, starts_with("Sepal")) |>  
8       head(3)  
9   } else if (indicateur == "Pétale") {  
10    resultat <- iris |>  
11      select(Species, starts_with("Petal")) |>  
12      head(3)  
13  }  
14  return(resultat)  
15  
16 }
```

3.3 La fonction browser (3/3)

👍 Avantages du browser :

- Facile à mettre en place
- Permet l'exécution pas à pas
- Fonctionne dans les chunks quarto

👎 Inconvénients du browser :

- Intrusif dans le code
- Moins visuels que les breakpoints
- On oublie souvent de les retirer du code une fois les bugs résolus.



4 Les fonctions debug() / debugonce()



4.1 Les fonctions debug() / debugonce() (1/4)

Les fonctions debug() et debugonce() permettent d'insérer la fonction browser() via l'appel à une fonction.

- debug(fonction) : active le mode pas-à-pas pour la fonction à chaque appel (nécessite la fonction undebbug() pour le désactiver).
- debugonce(f) : active le mode debug uniquement au prochain appel de la fonction.



4.2 Les fonctions debug() / debugonce() (2/4)

```
1 exemple1 <- function(indicateur) {  
2  
3   if (indicateur == "Sepale") {  
4     resultat <- iris |>  
5       select(Species, starts_with("Sepal")) |>  
6       head(3)  
7   } else if (indicateur == "Petale") {  
8     resultat <- iris |>  
9       select(Species, starts_with("Petal")) |>  
10      head(3)  
11   }  
12   return(resultat)  
13  
14 }
```



4.3 Les fonctions debug() / debugonce() (3/4)

```
1 debug(exemple1) # pour lancer le debugueur sur la fonction
2 exemple1("petale") # maintenant le debugueur se lance qd on soumet la fonction
3 undebug(exemple1) # pour arrêter le debugueur sur la fonction
4
5 debugonce(exemple1) # Pour lancer le debugueur qu'une seule fois
6 exemple1("petale") # le debugueur ne va se lancer qu'une seule fois
```



4.4 Les fonctions debug() / debugonce() (4/4)

👍 Avantages des debug() :

- Facile à mettre en place
- Permet l'exécution pas à pas
- Ne modifie pas le code source des fonctions

👎 Inconvénients des debug() :

- Ne fonctionne qu'avec les fonctions nommées
- On oublie souvent de les retirer du code une fois les bugs résolus au moment des appels des fonctions.



5 Exercice 1



5.1 Énoncé (partie 1)

Ex. 1

- Copier-coller le code R suivant et résoudre les erreurs avec votre outil préféré :

```
1 library(readr)
2 library(dplyr)
3
4 ### Import des fichiers
5 # - COG 2025
6 COG_2025_04_01 <- read_delim("~/CERISE/02-Espace-de-Production/000_Referentiels/0040_Geo/COG/2025/COG_2025_04_01.csv",
7                               delim = ";", escape_double = FALSE, trim_ws = TRUE)
8
9 # - Nomenclature des régions
10 regions <- readRDS("donnees/regions.rds")
11
12 # Définition des régions de l'utilisateur
13 regions_utilisateur <- c("93","94")
14
15 RESULTAT <- COG_2025_04_01 |>
16   filter(REG %in% regions_utilisateur) |>
17   left_join(regions, by = c("REG" = "Code_region")) |>
18   summarise(
19     nb_communes = n(),
20     .by = c(libelle_region,dep)) |>
21   arrange(desc(libelle_region),desc(nb_communes))
```

5.2 Énoncé (partie 2)

Ex. 1

Voici la fonction `calculs()` à mettre dans un fichier R `fonctions.R`.

```
1 #' @name calculs
2 #'
3 #' @param niveau_zoom indique le niveau de zoom souhaité par l'utilisateur
4 #'
5 #' @return un dataframe avec différentes statistiques
6 calculs <- function(niveau_zoom) {
7
8   if (niveau_zoom %in% regions_utilisateur) {
9     RESULTAT_ZOOM <- RESULTAT |>
10     filter(REG == zoom) |>
11     summarise(
12       region = first(LIB_REG),
13       nb_dep = n(),
14       total_communes = sum(NB_COMMUNES),
15       moyenne_par_dep = mean(NB_COMMUNES),
16       min_dep = min(NB_COMMUNES),
17       max_dep = max(NB_COMMUNES)
18     )
19     return(RESULTAT)
20   } else {
21     return(print("Le zoom n'est pas disponible dans les données"))
22   }
23
24 }
```

5.3 Énoncé (partie 3)

Ex. 1

Suite de l'exercice et appel de la fonction `calculs()` :

```
1 # Choix du zoom
2 zoom <- "94"
3
4 # Compilation des fonctions
5 source("fonctions.R")
6
7 # Lancement des calculs
8 calculs(zoom)
```



6 L'option recover



6.1 L'option recover (1/4)

L'option `options(error = recover)` sert à déboguer une erreur quand elle se produit.

Quand une erreur se produit, cela :

- interrompt l'exécution
- affiche la pile des fonctions appelées
- permet de choisir un niveau (=frame) pour inspecter les variables à ce moment-là.

`options(error = NULL)` permet de rétablir le comportement par défaut.

6.2 L'option recover (2/4)

```
1 options(error = recover)
2
3 f <- function(x) {
4   y <- x + 1
5   g(y)
6 }
7
8 g <- function(z) {
9   if (z < 10) {
10     res <- z*z
11   }
12   return(res)
13 }
```



6.3 L'option recover (3/4)

```
1 f(2) # le mode debug ne s'active pas
2 f(100) # le mode debug s'active
```

On obtient dans la console :

```
1 > f(100) # le mode debug s'active
2 Erreur dans g(y) : objet 'res' introuvable
3
4 Enter a frame number, or 0 to exit
5
6 1: f(100)
7 2: #3: g(y)
8
9 Sélection :
```

=> Taper 0, 1 ou 2 dans ce cas pour lancer le mode debug au niveau souhaité.

6.4 L'option recover (4/4)

👍 Avantages de l'option recover :

- Facile à mettre en place
- Permet rapidement de voir toutes les fonctions imbriquées qui ont conduit à l'erreur
- Ne modifie pas le code source des fonctions

👎 Inconvénients de l'option recover :

- Le mode debug ne s'active qu'après une erreur
- Pas utilisable pour les batchs par exemple

7 Astuce dans les chaine de traitements



7.1 Une astuce (1/5)

A l'intérieur d'une chaîne de traitements {tidyverse} avec %>%, il est possible d'afficher ou de sauvegarder un résultat intermédiaire pour résoudre rapidement un bug.

Si on veut simplement afficher le résultat intermédiaire, la syntaxe générale est :

```
1 data %>%  
2 etape_une() %>%  
3 { print(head(.)); . } %>%  
4 etape_deux()
```



7.2 Une astuce (2/5)

Un exemple :

```
1 iris %>%  
2   group_by(Species) %>%  
3   summarise(moyenne_long_petale = mean(Petal.Length, na.rm = TRUE)) %>%  
4   { print(head(.)); . } %>%  
5   rename(Especes = Species)
```

On obtient dans la console :

```
1 # A tibble: 3 × 2  
2   Species      moyenne_long_petale  
3   <fct>          <dbl>  
4 1 setosa          1.46  
5 2 versicolor     4.26  
6 3 virginica       5.55
```

7.3 Une astuce (3/5)

Si on veut stocker le résultat intermédiaire, , la syntaxe générale est :

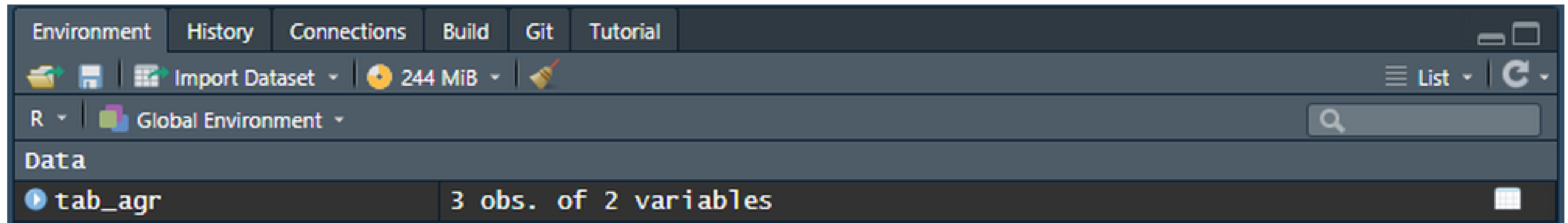
```
1 data %>%  
2 etape_un() %>%  
3 { result_interm <- . } %>% # On peut donner le nom que l'on veut au résultat intermédiaire  
4 etape_deux()
```



7.4 Une astuce (5/5)

```
1 iris %>%  
2   group_by(Species) %>%  
3   summarise(moyenne_long_petale = mean(Petal.Length, na.rm = TRUE)) %>%  
4   { tab_agr <- . } %>%  
5   rename(Especes = Species)
```

On obtient dans l'environnement :



7.5 Une astuce (5/5)

👍 Avantages de l'astuce :

- Pas d'interruption du code
- Pratique à mettre en place pour un debug rapide

👎 Inconvénients de l'astuce :

- Ne fonctionne qu'avec `%>%` (pas avec `|>`)
- On oublie souvent de les retirer du code une fois les bugs résolus



8 Le package boomer



8.1 Le package boomer

Le package {boomer} fournit des outils de débogage qui permettent d'inspecter les résultats intermédiaires d'un code R.

L'output produit par le package se présente sous la forme de bombes 💣 d'où le nom.

Ses 2 principales fonctions sont :

- `boom()` qui permet de diagnostiquer rapidement la performance d'un bloc de code R.
- `rig()` qui permet de tracer une fonction R et affiche automatiquement les mesures de performances à chacun de ses appels.

8.2 La fonction boom() (1/10)

3 écritures sont possibles :

- Ecriture en une ligne

```
1  
2 boom(iris |> head(2) |> filter(Sepal.Length > 5))
```

- Ecriture avec des crochets

```
1 boom({  
2   iris |>  
3     head(2) |>  
4     filter(Sepal.Length > 5)  
5 })
```

8.3 La fonction boom() (2/10)

- Ecriture à la fin de la chaîne de traitement (préférable)

```
1 iris |>  
2   head(2) |>  
3   filter(Sepal.Length > 5) |>  
4   boom()
```



8.4 La fonction boom() (3/10)

Le résultat dans la console :







```
1  💣 filter(head(iris, 2), Sepal.Length > 5)
2  · 💣 💣 head(iris, 2)
3  · Sepal.Length Sepal.Width Petal.Length Petal.Width Species
4  · 1           5.1         3.5         1.4         0.2    setosa
5  · 2           4.9         3.0         1.4         0.2    setosa
6  ·
7  · 💣 💣 Sepal.Length > 5
8  · [1]  TRUE  FALSE
9  ·
10 💣 filter(head(iris, 2), Sepal.Length > 5)
11 Sepal.Length Sepal.Width Petal.Length Petal.Width Species
12 1           5.1         3.5         1.4         0.2    setosa
13
14 Sepal.Length Sepal.Width Petal.Length Petal.Width Species
15 1           5.1         3.5         1.4         0.2    setosa
```

8.5 La fonction boom() (4/10)

En cas d'erreur, la fonction `boom()` affiche les résultats intermédiaires jusqu'à l'apparition du bug.

```
1 iris |>
2   head(2) |>
3   filter(Species == "virginica") |>
4   boom()
```

renvoie la sortie suivante :

```
1  filter(head(iris, 2), Species == "virginica")
2 ·   head(iris, 2)
3 ·   Sepal.Length Sepal.Width Petal.Length Petal.Width Species
4 · 1           5.1           3.5           1.4           0.2 setosa
5 · 2           4.9           3.0           1.4           0.2 setosa
6 ·
7 ·   Species == "virginica"
8 · [1] FALSE FALSE
9 ·
10  filter(head(iris, 2), Species == "virginica")
11 [1] Sepal.Length Sepal.Width Petal.Length Petal.Width Species
12 <0 lignes> (ou 'row.names' de longueur nulle)
13
14 [1] Sepal.Length Sepal.Width Petal.Length Petal.Width Species
15 <0 lignes> (ou 'row.names' de longueur nulle)
```

8.6 La fonction boom() (5/10)

La fonction `boom()` comprend 2 arguments très utiles :

- `clock` : booléen qui permet de quantifier le temps d'exécution de chaque étape.
- `print` : qui permet de fixer la manière dont sont affichés les résultats intermédiaires.
 - Soit `str`
 - Soit `glimpse`
 - Soit `list(data.frame = str)`

8.7 La fonction boom() (6/10)

Un exemple d'utilisation de `boom()` avec l'argument `clock = TRUE`


```
1 iris |>
2   head(2) |>
3   filter(Sepal.Length > 5) |>
4   boom(clock = TRUE)
```

Renvoie (partie 1) :

```
1 💣 filter
2 · 💣💣 head(iris, 2)
3 time: 0.215 ms
4 ·   Sepal.Length Sepal.Width Petal.Length Petal.Width Species
5 · 1           5.1         3.5         1.4         0.2   setosa
6 · 2           4.9         3.0         1.4         0.2   setosa
7 ·
8 · 💣💣 Sepal.Length > 5
9 time: 0.027 ms
10 · [1]  TRUE  FALSE
11 ·
```

8.8 La fonction boom() (7/10)

Renvoie (partie 2) :

```
1  filter(head(iris, 2), Sepal.Length > 5)
2 time: 0.001 s
3   Sepal.Length Sepal.Width Petal.Length Petal.Width Species
4 1           5.1         3.5         1.4         0.2   setosa
5
6   Sepal.Length Sepal.Width Petal.Length Petal.Width Species
7 1           5.1         3.5         1.4         0.2   setosa
8 >
```

8.9 La fonction boom() (8/10)

Un exemple d'utilisation de `boom()` avec l'argument `print = list(data.frame = str)`


```
1 iris |>
2   head(2) |>
3   filter(Sepal.Length > 5) |>
4   boom(clock = TRUE, print = list(data.frame = str))
```

Renvoie (partie 1) :

```
1  💣 filter
2  · 💣 💣 head(iris, 2)
3  time: 0.214 ms
4  · 'data.frame': 2 obs. of 5 variables:
5  · $ Sepal.Length: num 5.1 4.9
6  · $ Sepal.Width : num 3.5 3
7  · $ Petal.Length: num 1.4 1.4
8  · $ Petal.Width : num 0.2 0.2
9  · $ Species      : Factor w/ 3 levels "setosa","versicolor",...: 1 1
10 ·
11 · 💣 💣 Sepal.Length > 5
12 time: 0.021 ms
13 · [1] TRUE FALSE
14 ·
```

8.10 La fonction boom() (9/10)

Renvoie (partie 2) :

```
1  filter(head(iris, 2), Sepal.Length > 5)
2 time: 0.002 s
3 'data.frame': 1 obs. of 5 variables:
4 $ Sepal.Length: num 5.1
5 $ Sepal.Width : num 3.5
6 $ Petal.Length: num 1.4
7 $ Petal.Width : num 0.2
8 $ Species      : Factor w/ 3 levels "setosa","versicolor",...: 1
9
10 Sepal.Length Sepal.Width Petal.Length Petal.Width Species
11 1           5.1           3.5           1.4           0.2 setosa
```

8.11 La fonction boom() (10/10)

Quelques options utiles :

- Pour fixer l'option clock à TRUE tout le temps : `options(boomer.clock = TRUE)`
- Pour fixer l'option clock à TRUE tout le temps : `options(boomer.print = str)`
- Voir [ici](#) pour les options disponibles...

Pour aller plus loin :

- {boomer} contient les fonctions `boom_on()` et `boom_off()` qui peuvent être utilisées dans le mode debug pas à pas de RStudio (voir [cette page](#)).
- Le package contient également une fonctionnalité `boom_shinyApp()` (encore expérimentale) pour aider les développeurs à debugger les applications shiny (voir [cette page](#)).

8.12 La fonction `rig()` (1/3)

La fonction `rig()`

Soit la fonction suivante :

```
1 mafonction <- function(x) {  
2   y <- x^2  
3   return(y)  
4 }
```



8.13 La fonction `rig()` (2/3)

On appelle la fonction avec `rig()` :

```
1 rig(mafonction)(2)
```

Et on obtient :

```
1 🖱️ mafonction
2 x :
3 [1] 2
4 💣💣 x^2
5 [1] 4
6
7 🖱️ mafonction
8 [1] 4
```

Pour arrêter de surveiller une fonction, il existe la fonction `unrig()`.

8.14 La fonction rig() (3/3)

Avec une erreur :

```
1 rig(mafonction)("2")
```

Renvoie :

```
1 📌 mafonction
2 x :
3 [1] "2"
4 💣💣 x^2
5 Error: simpleError/error/condition
6 Erreur dans .Primitive("^")(x, 2) :
7   argument non numérique pour un opérateur binaire
8 📌 mafonction
```


8.15 Avantages/Inconvénients de {boomer}

👍 Avantages de {boomer} :

- Permet de “décortiquer” facilement les blocs de code (avec `boom()`)
- Permet de “décortiquer” facilement les fonctions (avec `rig()`)
- Peut aider à debugger les applications Shiny

👎 Inconvénients de {boomer} :

- Pour le debogage d’une longue chaine de traitements, le pas à pas est plus pratique.
- Surcharge le code source
- On oublie souvent de les retirer du code une fois les bugs résolus.



9 Exercice 2



Astuce

- ```
1 options(boomer.clock = TRUE)
2
3 library(boomer)
4
5 # - Nomenclature des régions
6 regions <- readRDS("donnees/regions.rds")
7
8 resultat <- readRDS("/var/data/nfs/CERISE/00-Espace-Personnel/damien.dotta/CERISE/03-Espace-de-Diffusion/100_Comptes_Revenus/10010_RICA/10_Donnees_definitives/RICA_10010_Donnees_definitives.rds")
9 left_join(regions,
10 by = c("REGIO" = "code_region")) %>%
11 filter(REGIO == "11") %>%
12 select(IDENT, SIRET, REGIO) %>%
13 boom()
```

# 10 Les logs



## 10.1 Les logs

Pour compléter les différents outils de débogage, il reste une pratique essentielle à présenter qui est **l'écriture des logs**.

Les logs permettent de tracer l'historique des actions, de suivre le cheminement logique du code R et ainsi de repérer l'origine des bugs/erreurs.

Cette pratique est tellement répandue et utile qu'il existe de très nombreux packages R : `{logger}`, `{futile.logger}`, `{logging}`, `{lgr}`, `{loggit}`, `{log4r}` ...

Dans cette présentation, nous allons utiliser le package `{logr}`.

## 10.2 Pourquoi logr ?

Le package {logr} présente plusieurs avantages :

- Il est très simple à utiliser
- Il fournit automatique un en-tête complet très utile au débogage
- L'horodatage est automatique (utile pour des mesures de performances)



## 10.3 logr (1/15)

Avec {logr}, il y a besoin de seulement 3 étapes pour créer un fichier de log :

- Ouvrir le fichier de log (avec `log_open()`)
  - Syntaxe : `log_open(chemin_vers_fichier_log)`
- Ecrire dans le fichier de log (avec `log_print()`)
  - Syntaxe : `log_print("message à écrire dans la log")`
- Fermer le fichier de log (avec `log_close()`)
  - Syntaxe : `log_close()`

## 10.4 logr (2/15)

### Structure d'une log :

- **Un en-tête** qui contient des informations essentielles : le chemin du fichier de log, le répertoire de travail, la version de R Core, le nom de l'utilisateur, le time-stamp d'exécution, les packages utilisés...
- **Un corps/body** qui constitue le coeur de la log. Il contient tous les messages et les objets que vous voulez afficher tels qu'il s'afficherait dans la console de R.
- **Un pied de page** qui contient le timestamp de fin d'exécution ainsi que le temps total écoulé pendant l'exécution du script R.



## 10.5 logr (3/15)

Un exemple :

```
1 library(logr)
2
3 # Ouverture de la log
4 log_open("ma_log.log")
5
6 log_print("## Lecture des données")
7 donnees <- readRDS("data/iris.rds")
8
9 log_print("## Traitement des données")
10 resultat <- donnees |>
11 group_by(Species) |>
12 summarise(moy_long_petales = mean(Petal.Length, na.rm = TRUE))
13
14 # Fermeture de la log
15 log_close()
```

# 10.6 logr (4/15)

Renvoie la log suivante (partie 1) :

```
1 =====
2 Log Path: ./log/ma_log.log
3 Working Directory: /var/data/nfs/CERISE/00-Espace-Personnel/damien.dotta
4 User Name: damien.dotta
5 R Version: 4.4.1 (.2024-06-14)
6 Machine: stats-prod-rstudio-2.zsg.cdpagri.fr x86_64
7 Operating System: Linux 4.18.0-372.32.1.el8_6.x86_64 #1 SMP Fri Oct 7 12:35:10 EDT 2022
8 Base Packages: stats graphics grDevices utils datasets methods base
9 Other Packages: dplyr_1.1.4 logr_1.3.8 common_1.1.3
10 Log Start Time: 2025-06-27 14:18:06.072628
11 =====
```

# 10.7 logr (5/15)

Partie 2 :

```
1 ## Lecture des données
2
3 NOTE: Log Print Time: 2025-06-27 14:18:06.08348
4 NOTE: Elapsed Time: 0.00261092185974121 secs
5
6 ## Traitement des données
7
8 NOTE: Log Print Time: 2025-06-27 14:18:06.090004
9 NOTE: Elapsed Time: 0.00652408599853516 secs
10
11 =====
12 Log End Time: 2025-06-27 14:18:06.120677
13 Log Elapsed Time: 0 00:00:00
14 =====
```

## 10.8 logr (6/15)

La fonction `log_print()` a également un alias utile pour la présentation des logs :

- `sep()` ajoute du texte dans la log comme `log_print()` mais ajoute des séparateurs avant/après le message.  
Cet exemple :

```
1 sep("Création des graphiques")
```

Va générer dans la log :

```
1 =====
2 Création des graphiques
3 =====
```

- `put()` est également un alias plus court que le nom `log_print()` dont le comportement est identique.

## 10.9 logr (7/15)

La fonction `log_print()` (ou `put()`) permet également d'envoyer des résultats dans les logs.

Par exemple :

```
1 log_print("## Traitement des données")
2 resultat <- donnees |>
3 group_by(Species) |>
4 summarise(moy_long_petales = mean(Petal.Length, na.rm = TRUE)) %>%
5 log_print()
```

# 10.10 logr (8/15)

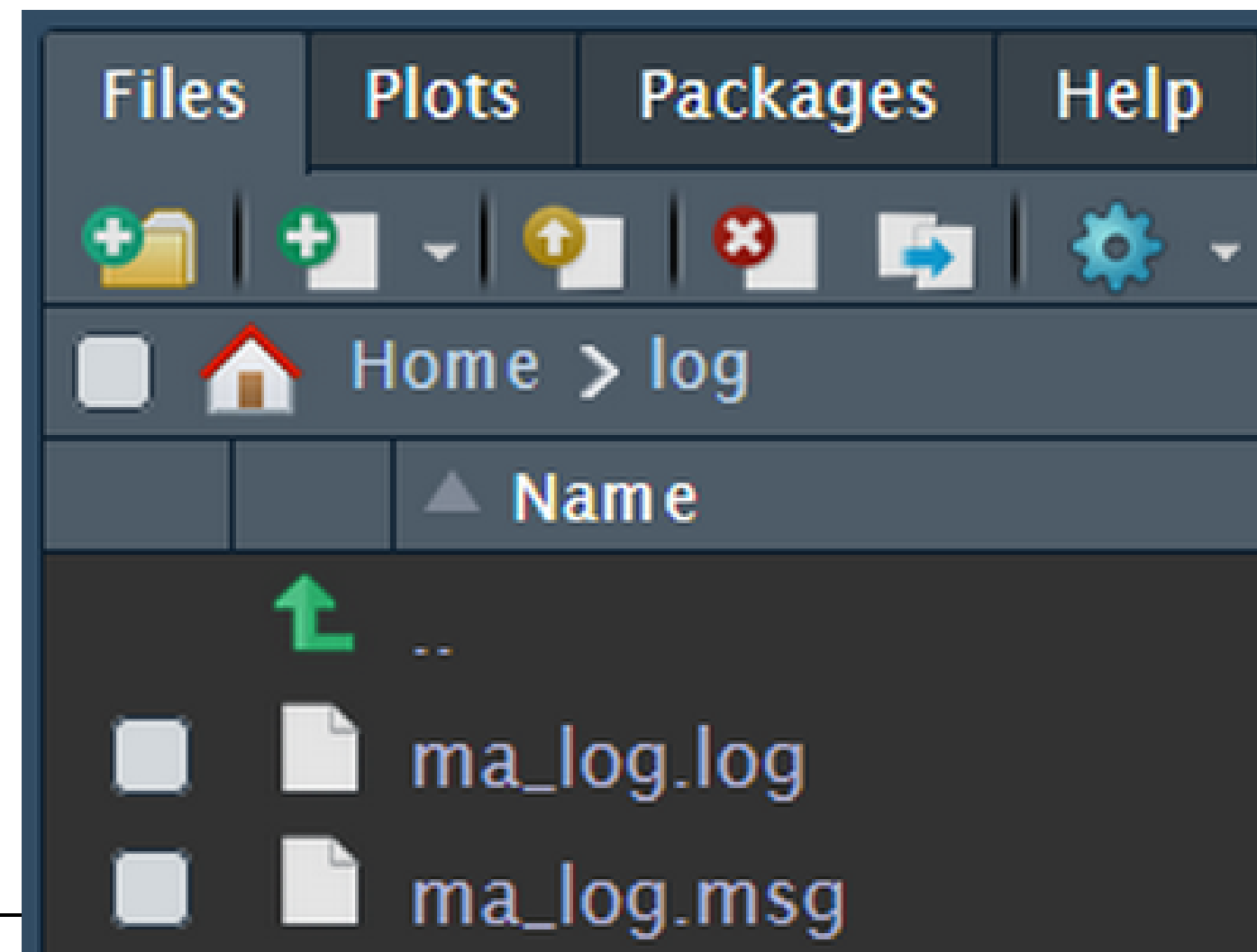
Renvoi :

```
1 ## Traitement des données
2
3 NOTE: Log Print Time: 2025-06-27 14:35:15.474827
4 NOTE: Elapsed Time: 0.0325253009796143 secs
5
6 # A tibble: 3 × 2
7 Species moy_long_petales
8 <fct> <dbl>
9 1 setosa 1.46
10 2 versicolor 4.26
11 3 virginica 5.55
12
13 NOTE: Data frame has 3 rows and 2 columns.
14
15 NOTE: Log Print Time: 2025-06-27 14:35:15.572858
16 NOTE: Elapsed Time: 0.0980312824249268 secs
```

## 10.11 logr (9/15)

En cas d'erreur/bug (ou de warnings) pendant l'exécution du programme R, ceux-ci s'afficheront dans la log à l'endroit où ils ont eu lieu.

En complément, {logr} va isoler les messages d'erreurs et les warnings dans un fichier qui porte l'extension `.msg`.



## 10.12 logr (10/15)

Le traceback de R peut être très verbeux en cas d'erreurs. Si vous n'avez pas besoin de ce niveau de précision, il est possible de le désactiver ponctuellement avec `log_print("traceback" = FALSE)` ou globalement avec `options("logr.traceback" = FALSE)`.

Exemple d'erreur dans les logs :

```
1 Error in group_by(.donnees, species) : [38;5;255m [31m*[38;5;255m Column `species` is not found.[39m
2 Traceback:
3 [1] "10: stop(fallback)"
4 [2] "9: signal_abort(cnd, .file)"
5 [3] "8: abort(bullets, call = error_call)"
6 [4] "7: group_by_prepare(.data, ..., .add = .add, error_call = current_env())"
7 [5] "6: group_by.data.frame(donnees, species)"
8 [6] "5: group_by(donnees, species)"
9 [7] "4: summarise(group_by(donnees, species), moy_long_petales = mean(Petal.Length, "
10 [8] " na.rm = TRUE))"
11 ...
```

Sans traceback :

```
1 Error in group_by(.donnees, species) : [38;5;255m [31m*[38;5;255m Column `species` is not found.[39m
```



## 10.13 logr (11/15)

Quelques conseils si vous savez que vous allez utiliser logr sur votre programme :

- Remplacer vos commentaires dans votre code R par des `log_print()`

A la place de :

```
1 # Traitement des données
```

Mettre ça :

```
1 log_print("# Traitement des données")
```

- A chaque exécution de vos programmes R, les logs seront par défaut écrasées. Si vous souhaitez les conserver, il est possible d'horodater leurs noms avec par exemple :

```
1 nom <- paste0("mylog_", format(Sys.Date(), "%Y-%m-%d"))
2
3 # Open log
4 lf <- log_open(nom)
```

## 10.14 logr (12/15)

Il est possible de logger seulement une partie d'un script R. Pour cela, il existe les options `options("logr.on" = FALSE)` et `options("logr.on" = TRUE)`.

```
1 # Un 1er traitement
2 iris |> names()
3
4 # Activation de logr
5 options("logr.on" = TRUE)
6
7 log_open("ma_log.log")
8
9 log_print("## Lecture des données")
10 donnees <- readRDS("data/iris.rds")
11
12 log_close()
13
14 # Désactivation de logr
15 options("logr.on" = FALSE)
16
17 # La suite du programme
18 donnees |> ...
```

## 10.15 logr (13/15)

- Si vous souhaitez raccourcir la longueur de vos logs, il est possible de supprimer les lignes blanches avec l'option `options("logr.compact" = TRUE)`.
- Si vous souhaitez ne pas écrire les notes dans vos logs, il existe l'option `options("logr.notes" = FALSE)`
- Si vous souhaitez envoyer vos variables d'environnement dans vos logs, c'est possible via `put(Sys.getenv())`
- Pour plus de conseils, consultez [la FAQ du package](#) qui est très riche.

## 10.16 logr (14/15)

Si votre code est essentiellement écrit avec {dplyr} et {tidyr}, il est possible de limiter le nombre d'appels à `log_print()` en utilisant `options("logr.autolog" = TRUE)`.

Cela permet d'écrire dans les logs automatiquement (grâce à `tidylog`) à chaque appel d'un bloc de code.

Exemple sans `log_print()` :

```
1 options("logr.autolog" = TRUE)
2
3 log_open("ma_log.log")
4
5 donnees <- readRDS("data/iris.rds")
6
7 resultat <- donnees |>
8 group_by(Species) |>
9 summarise(moy_long_petales = mean(Petal.Length, na.rm = TRUE))
10
11 # Fermeture de la log
12 log_close()
```

# 10.17 logr (15/15)

Renvoie la log suivante :

```
1 En-tête...
2 =====
3
4 group_by: one grouping variable (.Species)
5
6 NOTE: Log Print Time: 2025-06-27 15:27:52.030858
7 NOTE: Elapsed Time: 0.0213210582733154 secs
8
9 summarise: now 3 rows and 2 columns, ungrouped
10
11 NOTE: Log Print Time: 2025-06-27 15:27:52.346752
12 NOTE: Elapsed Time: 0.315893888473511 secs
13
14 =====
15 Pied de page..
```

# 11 Exercice 3



# 11.1 Énoncé (partie 1)

## Astuce

- Modifier le code R suivant afin d'obtenir la log souhaitée avec {logr}.

```
1 Sys.getenv("http_proxy")
2
3 set.seed(123)
4 n <- 10000000
5 species <- c("setosa", "versicolor", "virginica")
6
7 simu_data <- data.frame(
8 Species = sample(species, n, replace = TRUE),
9 Sepal.Length = rnorm(n, mean = 6, sd = 0.8),
10 Sepal.Width = rnorm(n, mean = 3, sd = 0.5),
11 Petal.Length = rnorm(n, mean = 4, sd = 1.2),
12 Petal.Width = rnorm(n, mean = 1.2, sd = 0.4)
13)
14
15 # Calculer les moyennes et écarts-types
16 resultats <- simu_data %>%
17 group_by(Species) %>%
18 summarise(
19 mean_sepal_length = mean(Sepal.Length),
20 sd_sepal_length = sd(Sepal.Length),
21 mean_petal_length = mean(Petal.Length),
22 sd_petal_length = sd(Petal.Length),
23 n = n()
24)
```

## 11.2 Énoncé (partie 2)

### Astuce

```
1 # Prédire Petal.Length en fonction de Sepal.Length
2 modele <- lm(Petal.Length ~ Sepal.Length, data = simu_data)
3
4
5 # Affichage d'un sous-échantillon pour la clarté du graphique
6 simu_data_sample <- simu_data %>% sample_n(5000)
7
8 # Création du graphique
9 ggplot(simu_data_sample, aes(x = Sepal.Length, y = Petal.Length)) +
10 geom_point(alpha = 0.2, color = "steelblue") +
11 geom_smooth(method = "lm", color = "darkred", se = FALSE) +
12 labs(
13 title = "Régression linéaire : Petal.Length ~ Sepal.Length",
14 x = "Longueur du sépale",
15 y = "Longueur du pétale"
16) +
17 theme_minimal()
18
19
20 # Résumé statistique du modèle
21 summary(modele)
```



# 11.3 Log souhaitée (en-tête)

Lien vers [la log complète](#).

Astuce

```
1 =====
2 Log Path: ./log/log_exo3.log
3 Program Path: /var/data/nfs/CERISE/00-Espace-Personnel/damien.dotta/test_log.R
4 Working Directory: /var/data/nfs/CERISE/00-Espace-Personnel/damien.dotta
5 User Name: damien.dotta
6 R Version: 4.4.1 (2024-06-14)
7 Machine: stats-prod-rstudio-2.zsg.cdpagri.fr x86_64
8 Operating System: Linux 4.18.0-372.32.1.el8_6.x86_64 #1 SMP Fri Oct 7 12:35:10 EDT 2022
9 Base Packages: stats graphics grDevices utils datasets methods base
10 Other Packages: tidylog_1.1.0 ggplot2_3.5.1 dplyr_1.1.4 logr_1.3.8 common_1.1.3
11 Log Start Time: 2025-07-07 16:08:38.451985
12 =====
```



# 11.4 Log souhaitée (body - partie 1)

## Astuce

```

1 Affichage var. d'environnement
2
3 http://rie.proxy.national.agri:8080/
4
5 =====
6 Création jeu de données
7 =====
8
9 Aperçu des données
10
11 Species Sepal.Length Sepal.Width Petal.Length Petal.Width
12 1 virginica 6.271247 2.846290 5.953113 1.8597240
13 2 virginica 5.470469 3.518637 4.072064 1.3026571
14 3 virginica 7.320498 3.167360 3.158868 1.6208414
15 4 versicolor 5.408678 3.246124 4.136058 0.8419174
16 5 virginica 5.968974 2.366812 2.806698 0.8434953
17 6 versicolor 4.926451 3.220669 1.126885 1.9263393

```

# 11.5 Log souhaitée (body - partie 2)

## Astuce

```
1 =====
2 Traitement des données
3 =====
4
5 group_by: one grouping variable .(Species)
6
7 summarise: now 3 rows and 6 columns, ungrouped
8
9 =====
10 Analyse de données
11 =====
12
13 2025-07-07 16:08:44# Prédire Petal.Length en fonction de Sepal.Length
14
15 2025-07-07 16:08:48# Graphique sur un sous-échantillon
16
17 sample_n: removed 9,995,000 rows .(>99%), 5,000 rows remaining
```



# 11.6 Log souhaitée (body - partie 3)

## Astuce

```
1 =====
2 A retenir
3 =====
4 Call:
5 lm(formula = Petal.Length ~ Sepal.Length, data = simu_data)
6
7 Residuals:
8 Min 1Q Median 3Q Max
9 -6.2654 -0.8094 0.0001 0.8095 6.3613
10
11 Coefficients:
12 Estimate Std. Error t value Pr(>|t|)
13 (Intercept) 3.9968280 0.0028715 1391.89 <2e-16 ***
14 Sepal.Length 0.0003701 0.0004744 0.78 0.435
15 ---
16 Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
17
18 Residual standard error: 1.2 on 9999998 degrees of freedom
19 Multiple R-squared: 6.087e-08, Adjusted R-squared: -3.913e-08
20 F-statistic: 0.6087 on 1 and 9999998 DF, p-value: 0.4353
```

# 11.7 Log souhaitée (pied de page)

## Astuce

```
1 =====
2 Log End Time: 2025-07-07 16:09:00.851627
3 Log Elapsed Time: 0 00:00:22
4 =====
```



# 12 Bibliographie



## 12.1 Liens complémentaires

Pour en savoir plus :

- [Debugging with the RStudio IDE](#)
- [Debugging - Advanced R](#)
- [Package boomer](#)
- [Package logr](#)