## Introduction à R

The R Base Package

INESSS, BDCA

2025-04-25

# Les bases

# Arithmétique (1)

```
# Addition
5 + 5
#> [1] 10
# Soustraction
5 - 5
#> [1] 0
# Multiplication
3 * 5
#> [1] 15
# Division
(5 + 5) / 2
#> [1] 5
```

## Arithmétique (2)

```
# Exposant
2^5

#> [1] 32

# Modulo
28 %% 6

#> [1] 4
```

Prendre note que le caractère # met la ligne de code en commentaire et n'est pas évalué par R.

#### Affectation de variables (1)

Une variable permet de stocker une valeur ou un objet. On peut ensuite l'utiliser ultérieurement pour accéder facilement à cette valeur.

```
# Assigner la valeur 42 à x
x <- 42

# Afficher/Imprimer la valeur de la variable x
x

#> [1] 42

print(x)
#> [1] 42
```

#### Affectation de variables (2)

Nous pouvons appliquer des opérations arithmétiques entre des variables numériques.

```
x <- 5
y <- 25
x + y

#> [1] 30

x - y

#> [1] -20

y / x

#> [1] 5
```

### Type de données de base (1)

• Nombres entiers: 4 (integer/int)

```
integer <- 4
```

• Numérique: 4.5 (numeric/num)

```
numeric <- 4.5
```

• Logique ou booléen : TRUE ou FALSE (logical/logi)

```
logique1 <- TRUE
logique2 <- FALSE
```

• Texte: Chaîne de caractères (character/chr)

```
character1 <- "Chaine de caractères"
character2 <- 'Chaine de caractères'
```

### Type de données de base (2)

R peut décider que c'est numérique même si on veut des nombres entiers. Ajouter L après un nombre force la classe d'un nombre à entier au lieu de numérique.

```
class(5)

#> [1] "numeric"

class(5L)

#> [1] "integer"

class(1:5)

#> [1] "integer"
```

# **Vecteurs**

#### Créer un vecteur

On crée un vecteur avec la fonction c().

```
numeric_vec <- c(1, 10, 49)
character_vec <- c("a", "b", "c")
boolean_vec <- c(TRUE, FALSE, T, F)</pre>
```

#### Mise en situation

Médicament ajouté ou déduit de l'inventaire à la fin de la journée du lundi au vendredi :

```
# Advil
advil <- c(140, -50, 20, -120, 240)

# Tylenol
tylenol <- c(-24, 50, 100, -350, 10)
```

#### Nommer un vecteur (1)

Nous pouvons donner un nom aux éléments d'un vecteur avec la fonction names ().

```
advil \leftarrow c(140, -50, 20, -120, 240)
advil
#> [1] 140 -50 20 -120 240
tylenol \langle -c(-24, 50, 100, -350, 10) \rangle
names(advil) <- c(</pre>
   "Lundi", "Mardi", "Mercredi", "Jeudi", "Vendredi"
names(tylenol) <-</pre>
  c("Lundi", "Mardi", "Mercredi", "Jeudi", "Vendredi")
print(advil)
                Mardi Mercredi
                                   Jeudi Vendredi
#>
      Lundi
#>
        140
                  -50
                             20
                                -120
                                                240
```

#### Nommer un vecteur (2)

Si une information est utilisé plus d'une fois, il est utile de créer une nouvelle variable.

```
advil <- c(140, -50, 20, -120, 240)
tylenol <- c(-24, 50, 100, -350, 10)

# Jour de la semaine
jours <- c("Lundi", "Mardi", "Mercredi", "Jeudi", "Vendredi")

# Attribuer le jour de la semaine
names(advil) <- jours
names(tylenol) <- jours
print(advil)</pre>
```

```
#> Lundi Mardi Mercredi Jeudi Vendredi
#> 140 -50 20 -120 240
```

#### Calculs sur un vecteur (1)

Pour chaque jour, calculons le nombre de médicaments qui ont été ajouté ou déduit de l'inventaire.

```
advil
#>
      Lundi
                Mardi Mercredi
                                    Jeudi Vendredi
#>
        140
                  -50
                             20
                                     -120
                                                240
tylenol
#>
                Mardi Mercredi
                                    Jeudi Vendredi
      Lundi
        -24
                   50
                            100
                                     -350
#>
                                                 10
 inventaire_jour <- advil + tylenol</pre>
 inventaire_jour
      Lundi
                Mardi Mercredi
                                    Jeudi Vendredi
#>
        116
#>
                    0
                            120
                                     -470
                                               250
```

### Calculs sur un vecteur (2)

Les opérations de plusieurs vecteurs se font à partir de la position des valeurs, pas des noms.

```
advil
#>
      Lundi
               Mardi Mercredi
                                  Jeudi Vendredi
        140
                 -50
                            20
                                   -120
#>
                                              240
tylenol2 <- tylenol[c(3,2,5,1,4)] # modifier l'ordre des jours
tylenol2
#> Mercredi
               Mardi Vendredi
                                  Lundi
                                            Jeudi
        100
                                    -24
                                             -350
#>
                  50
                            10
inventaire_jour2 <- advil + tylenol2</pre>
inventaire_jour2
      Lundi
               Mardi Mercredi
                                  Jeudi Vendredi
#>
                                   -144
#>
        240
                   0
                            30
                                             -110
```

Les Advil du lundi sont additionné aux Tylenol du mercredi, ce qui n'a pas de sens.

### Calculs sur un vecteur (3)

#### Calculons l'inventaire de la semaine

```
total_advil <- sum(advil)</pre>
total_advil
#> [1] 230
total_tylenol <- sum(tylenol)</pre>
total_tylenol
#> [1] -214
total_semaine <- sum(inventaire_jour)</pre>
total_semaine
#> [1] 16
```

#### Comparer les inventaires

Vérifions si les inventaires sont plus élevés en Advil qu'en Tylenol.

```
total_advil

#> [1] 230

total_tylenol

#> [1] -214

total_advil > total_tylenol

#> [1] TRUE
```

#### Sélection dans un vecteur (1)

Pour sélectionner une valeur dans un vecteur, nous utilisons les crochets [].

Utiliser les doubles crochets focalise sur la valeur perdant ainsi l'information du nom (si présent)

```
vecteur_sans_nom <- c(5, 4, 7)
vecteur_sans_nom[2]
#> [1] 4
vecteur_sans_nom[[2]]
```

Aucune différence entre crochets simples ou doubles

**#>** [1] 4

### Sélection dans un vecteur (2)

Il est possible de sélectionner plusieurs valeurs à l'aide d'un vecteur. Si nous voulons analyser les inventaires du lundi et du vendredi, nous pourrions utiliser le vecteur c(1, 5) à l'intérieur des crochets : advil[c(1, 5)]. Analysons les inventaires du mardi au jeudi :

```
advil[c(2, 3, 4)]
      Mardi Mercredi
#>
                         Jeudi
#>
        -50
                   20
                          -120
advil[2:4]
      Mardi Mercredi
                         Jeudi
#>
#>
        -50
                   20
                          -120
```

#### Sélection dans un vecteur (3)

Une autre manière de sélectionner des valeurs est d'utiliser les noms lorsqu'il y en a.

```
advil
#>
      Lundi
               Mardi Mercredi
                                 Jeudi Vendredi
#>
        140
                 -50
                           20
                                  -120
                                             240
advil[c("Mardi", "Jeudi")]
#> Mardi Jeudi
#> -50 -120
advil["Mercredi"]
#> Mercredi
         20
#>
```

#### Sélection par comparaison (1)

Il est aussi possible de sélectionner des valeurs par comparaison.

```
# Jour(s)s où l'inventaire Advil est positif
advil > 0
      Lundi
              Mardi Mercredi Jeudi Vendredi
#>
      TRUE
              FALSE TRUE
                              FALSE
                                          TRUE
#>
advil[advil > 0] # insérer la condition dans les crochets
      Lundi Mercredi Vendredi
#>
       140
                          240
#>
                  20
selection <- advil > 0 # enregistrer la condition sous une variable
 advil[selection] # utiliser la variable dans les crochets
      Lundi Mercredi Vendredi
#>
#>
       140
                  20
                          240
 # Jour(s)s où l'inventaire total est de zéro
 advil[inventaire_jour == 0]
#> Mardi
    -50
#>
```

#### Sélection par comparaison (2)

Liste des opérateurs de comparaison :

- <: plus petit
- >: plus grand
- <= : plus petit ou égal
- >= : plus grand ou égal
- ==: égal\*
- != : pas égal, différent

<sup>\*</sup> important de mettre deux ==, sinon assignation d'une variable comme <-

# Les data frames

#### Qu'est-ce qu'un data frame?

#> Valiant

Un *data frame* est un tableau comme on peut en avoir avec Excel. Chaque colonne est une variable d'un même type (numérique, caractère, logique...) et chaque ligne est une observation.

```
head(mtcars)
#>
                   mpg cyl disp hp drat wt qsec vs am gear carb
#> Mazda RX4
                  21.0
                           160 110 3.90 2.620 16.46
                                                               4
#> Mazda RX4 Wag
                  21.0
                            160 110 3.90 2.875 17.02 0 1
                                                               4
#> Datsun 710
               22.8
                                                               1
                         4 108 93 3.85 2.320 18.61 1 1
#> Hornet 4 Drive 21.4
                         6 258 110 3.08 3.215 19.44 1
                                                               1
#> Hornet Sportabout 18.7
                           360 175 3.15 3.440 17.02
```

225 105 2.76 3.460 20.22

Les fonctions head() et tail() permettent de voir les premières ou les dernières observations d'un ensemble de données.

18.1

#### Structure des données

La fonction str() est semblable à la fonction class(), mais affiche plus d'informations.

```
class(mtcars)
#> [1] "data.frame"
str(mtcars)
#> 'data.frame':
                  32 obs. of 11 variables:
   $ mpg : num 21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...
   $ cvl : num 6646868446 ...
   $ disp: num 160 160 108 258 360 ...
   $ hp : num
               110 110 93 110 175 105 245 62 95 123 ...
#> $ drat: num 3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ...
#> $ wt : num 2.62 2.88 2.32 3.21 3.44 ...
#> $ gsec: num 16.5 17 18.6 19.4 17 ...
#> $ vs : num
               0 0 1 1 0 1 0 1 1 1 ...
#> $ am : num 1 1 1 0 0 0 0 0 0 0 ...
#> $ gear: num 4 4 4 3 3 3 3 4 4 4 ...
#> $ carb: num 4 4 1 1 2 1 4 2 2 4 ...
```

Il est préférable d'utiliser str() pour une première exploration que de visualiser tout le tableau avec View().

mtcars est un data. frame de 32 observations et 11 colonnes, toutes les variables sont de type numériques (num) et les dix premières observations de chaque colonne sont affichées.

#### Créer un data.frame (1)

La fonction data.frame() permet de créer un data.frame. Chaque colonne est un vecteur et tous les vecteurs doivent contenir le même nombre de valeurs.

NA est pour *Not Available / Missing Values*. Pas besoin de préciser si NA est dans un vecteur numérique, logique ou caractère.

#### Créer un data.frame (2)

Chaque vecteur doivent contenir le même nombre de valeurs. Les colonnes sont dans le même ordre que les vecteurs inscrits dans la fonction data.frame().

```
mon_data <- data.frame(id, personne, main, gentillesse_sur_10, on_laime)</pre>
mon data
    id personne
                      main gentillesse_sur_10 on_laime
#> 1 1 Jean Airoldi droitier
                                            FALSE
          Ricardo
#> 2 2
                      <NA>
                                             TRUE
#> 3 George Lucas ambidextre
                                            FALSE
#> 4 4 FredK Le Roi gaucher
                                          TRUE
                                        10
#> 5 5 Dre Samson
                  droitier
                                       20
                                             TRUE
str(mon_data)
#> 'data.frame': 5 obs. of 5 variables:
#> $ id
         : int 12345
#> $ personne : chr "Jean Airoldi" "Ricardo" "George Lucas" "FredK Le
#> $ main
         : chr "droitier" NA "ambidextre" "gaucher" ...
$ on_laime
                   : logi FALSE TRUE FALSE TRUE TRUE
```

#### Sélection des éléments d'un *data.frame* (1)

Comme pour les vecteurs, on sélectionne les éléments d'un tableau avec les crochets []. À l'aide d'une virgule, on peut décider entre les lignes et les colonnes à afficher :

- mon\_data[1:2, ]: Les deux premières lignes, toutes les colonnes.
- mon\_data[1, 2]: **Valeur** de la cellule de la première ligne et la deuxième colonne.
- mon\_data[1:3, c(2:4, 6)]: Les trois premières lignes et les colonnes 2, 3, 4 et 6.

#### Sélection des éléments d'un *data.frame* (2)

```
mon_data
       personne
                       main gentillesse_sur_10 on_laime
    id
#> 1 1 Jean Airoldi
                  droitier
                                               FALSE
           Ricardo
                                               TRUE
                  <NA>
FALSE
#> 4 4 FredK Le Roi gaucher
                                          10 TRUE
#> 5 5 Dre Samson droitier
                                          20
                                               TRUE
mon_data[1, 2]
#> [1] "Jean Airoldi"
mon_data[1:3, 2:4]
                    main gentillesse_sur_10
#>
       personne
#> 1 Jean Airoldi droitier
        Ricardo
#> 2
                    <NA>
                                        2
#> 3 George Lucas ambidextre
                                        0
```

#### Sélection des éléments d'un *data.frame* (3)

On peut utiliser le nom des colonnes au lieu les nombres

```
mon_data
                       main gentillesse_sur_10 on_laime
#>
    id
       personne
#> 1 1 Jean Airoldi droitier
                                               FALSE
           Ricardo <NA>
                                               TRUE
FALSE
#> 4 4 FredK Le Roi gaucher
                                         10 TRUE
#> 5 5 Dre Samson droitier
                                               TRUE
                                         20
mon_data[2, "gentillesse_sur_10"]
#> [1] 2
mon_data[c(1, 4, 5), c("personne", "main")]
#>
       personne
                   main
#> 1 Jean Airoldi droitier
#> 4 FredK Le Roi gaucher
#> 5 Dre Samson droitier
```

#### Sélection d'une colonne uniquement

Indiquer uniquement la ou les colonnes après la virgule retournera toutes les lignes.

En plus des crochets, l'utilisation du symbole \$ est aussi possible.

```
mon_data[, 2]

#> [1] "Jean Airoldi" "Ricardo" "George Lucas" "FredK Le Roi" "Dre Samson"

mon_data[, "personne"]

#> [1] "Jean Airoldi" "Ricardo" "George Lucas" "FredK Le Roi" "Dre Samson"

mon_data$personne

#> [1] "Jean Airoldi" "Ricardo" "George Lucas" "FredK Le Roi" "Dre Samson"
```

### Sélection d'observation selon condition (1)

À l'aide d'un vecteur contenant des TRUE ou des FALSE, on peut sélectionner les données des gens qu'on aime.

```
mon_data[, c("personne", "on_laime")]
        personne on_laime
#>
#> 1 Jean Airoldi
                     FALSE
         Ricardo
#> 2
                     TRUE
#> 3 George Lucas
                    FALSE
#> 4 FredK Le Roi
                     TRUE
      Dre Samson
                  TRUE
#> 5
mon_data$on_laime # la colonne 'on_laime' sous la forme d'un vecteur
#> [1] FALSE TRUE FALSE TRUE TRUE
# Utiliser le vecteur mon_data$on_laime pour sélectionner les valeurs TRUE
# uniquement. Les FALSE sont exclus.
mon_data[mon_data$on_laime, ]
                        main gentillesse_sur_10 on_laime
           personne
            Ricardo
                         <NA>
                                                     TRUE
#> 4 4 FredK Le Roi gaucher
                                                     TRUE
                                              10
         Dre Samson droitier
                                                     TRUE
                                              20
```

## Sélection d'observation selon condition (2)

Même chose que précédemment, mais en utilisant la variable on\_laime

```
mon_data[mon_data$on_laime, ] # avec colonne du data
    id
           personne main gentillesse_sur_10 on_laime
#>
#> 2 2
            Ricardo
                        <NA>
                                                   TRUE
#> 4 4 FredK Le Roi gaucher
                                                   TRUE
                                            10
#> 5 5 Dre Samson droitier
                                                   TRUE
                                            20
variable_on_laime <- mon_data$on_laime</pre>
mon_data[variable_on_laime, ] # avec la variable
                    main gentillesse_sur_10 on_laime
#>
    id
           personne
            Ricardo
                                                   TRUE
#> 2 2
                        <NA>
#> 4 4 FredK Le Roi gaucher
                                            10
                                                   TRUE
#> 5 5 Dre Samson droitier
                                                   TRUE
                                            20
```

### Sélection d'observation selon condition (3)

Cherchons les gens qui ont une gentillesse d'au moins 5 sur 10.

#> 4 4 FredK Le Roi gaucher

Dre Samson droitier

#> 5

```
# Fn utilisant la colonne du data
mon_data$gentillesse_sur_10 > 5
#> [1] TRUE FALSE FALSE TRUE TRUE
mon_data[mon_data$gentillesse_sur_10 > 5, ]
    id
           personne main gentillesse_sur_10 on_laime
#>
#> 1 1 Jean Airoldi droitier
                                                   FALSE
#> 4 4 FredK Le Roi gaucher
                                             10
                                                    TRUE
#> 5 5 Dre Samson droitier
                                                    TRUE
                                             20
# En utilisant une variable
gentil5 <- mon_data$gentillesse_sur_10 > 5
mon_data[gentil5, ]
                        main gentillesse_sur_10 on_laime
    id
           personne
#> 1 1 Jean Airoldi droitier
                                                   FALSE
```

34 / 46

**TRUE** 

TRUE

10

20

#### Trier son data frame

La fonction order () indique le rang, la position de la donnée. On peut ensuite l'utiliser pour afficher les lignes dans l'ordre voulu.

```
mon_data
                           main gentillesse_sur_10 on_laime
            personne
     id
#> 1 1 Jean Airoldi
                       droitier
                                                       FALSE
#> 2 2
             Ricardo
                                                       TRUE
                           <NA>
#> 3 George Lucas ambidextre
                                                       FALSE
#> 4 4 FredK Le Roi
                        gaucher
                                                       TRUE
                                                 10
#> 5 5
          Dre Samson
                       droitier
                                                 20
                                                       TRUE
tri_gentil <- order(mon_data$gentillesse_sur_10)</pre>
tri_gentil
#> [1] 3 2 1 4 5
mon_data[tri_gentil,]
                           main gentillesse_sur_10 on_laime
     id
#>
            personne
#> 3 George Lucas ambidextre
                                                       FALSE
             Ricardo
                                                       TRUE
#> 2 2
                           <NA>
#> 1 1 Jean Airoldi
                       droitier
                                                       FALSE
#> 4 4 FredK Le Roi
                      gaucher
                                                       TRUE
                                                 10
                       droitier
          Dre Samson
                                                       TRUE
#> 5 5
                                                 20
```

# **Facteurs**

### Qu'est-ce qu'un facteur? (1)

Le terme facteur est un type de données pour stocker des variables catégoriques et être en mesure de les trier. Elles sont aussi plus efficace au niveau de la mémoire.

```
sex <- c("Femme", "Homme", "Femme", "Femme")
factor_sex <- factor(sex)
factor_sex</pre>
```

```
#> [1] Femme Homme Homme Femme Femme
#> Levels: Femme Homme
```

S'ils ne sont pas défini, les *Levels* sont en ordre alphabétique.

# Qu'est-ce qu'un facteur (2)

#### Comment trier une variable catégorique :

```
temperature <- c("Élevée", "Faible", "Élevée", "Faible", "Moyenne")
factor_temperature <- factor(
  temperature, # vecteur à utiliser
  order = TRUE, # si l'ordre a un concept de 'plus petit/grand que'
  levels = c("Faible", "Moyenne", "Élevée") # ordre souhaité
)
factor_temperature</pre>
```

```
#> [1] Élevée Faible Élevée Faible Moyenne
#> Levels: Faible < Moyenne < Élevée</pre>
```

#### Très utile avec les légendes de graphiques.

```
mois_annee <- c("Déc", "Avr", "Jan", "Mar")
factor_mois_annee <- factor(
    mois_annee,
    order = FALSE,  # les mois ont un ordre, mais ne sont pas plus petit/grand qu'un autre
    levels = c(
        "Jan", "Fév", "Mar", "Avr", "Mai", "Jun",
        "Jul", "Aou", "Sep", "Oct", "Nov", "Déc"
    )
)
sort(factor_mois_annee)</pre>
```

```
#> [1] Jan Mar Avr Déc
#> Levels: Jan Fév Mar Avr Mai Jun Jul Aou Sep Oct Nov Déc
```

# Liste

#### Qu'est-ce qu'une liste?

Une liste dans R permet de réunir une variété d'objets sous une même variable. Peut contenir des matrices, des vecteurs, des *data.frame* et même d'autres listes.

```
MaListe = list(element1, element2, ...)
```

#### Comment créer une liste (1)

22.8 4 108

```
vecteur <- 1:10
list1 \leftarrow list(x = "a", y = "b", z = "c")
datafr <- mtcars[1:3, 1:3]</pre>
ma_liste <- list(vecteur, list1, datafr)</pre>
ma_liste
#> [[1]]
#> [1] 1 2 3 4 5 6 7 8 9 10
#>
#> [[2]]
#> [[2]]$x
#> [1] "a"
#>
#> [[2]]$y
#> [1] "b"
#>
#> [[2]]$z
#> [1] "c"
#>
#>
#> [[3]]
                  mpg cyl disp
#>
#> Mazda RX4
                 21.0 6 160
#> Mazda RX4 Wag 21.0 6 160
#> Datsun 710
```

#### Comment créer une liste (2)

Nommer directement les éléments d'une liste

```
ma_liste <- list(vec = vecteur, lis = list1, df = datafr)</pre>
ma liste
#> $vec
#> [1] 1 2 3 4 5 6 7 8 9 10
#>
#> $lis
#> $lis$x
#> [1] "a"
#>
#> $lis$y
#> [1] "b"
#>
#> $lis$z
#> [1] "c"
#>
#>
#> $df
#>
                mpg cyl disp
#> Mazda RX4
                21.0 6 160
#> Mazda RX4 Wag 21.0 6 160
#> Datsun 710
                22.8 4 108
```

#### Comment créer une liste (3)

Nommer les éléments d'une liste après l'avoir créée

```
names(ma_liste) <- c("MonVecteur", "MaListe", "MonDf")</pre>
ma_liste
#> $MonVecteur
    [1] 1 2 3 4 5 6 7 8 9 10
#>
#> $MaListe
#> $MaListe$x
#> [1] "a"
#>
#> $MaListe$y
#> [1] "b"
#>
#> $MaListe$z
#> [1] "c"
#>
#>
#> $MonDf
#>
                 mpg cyl disp
#> Mazda RX4
                21.0 6 160
#> Mazda RX4 Wag 21.0 6 160
#> Datsun 710
                22.8 4 108
```

### Sélectionner les éléments d'une liste (1)

Semblable à la sélection d'un vecteur ou d'un *data frame* avec une particularité au niveau des crochets []

```
ma_liste[1]
#> $MonVecteur
#> [1] 1 2 3 4 5 6 7 8 9 10
str(ma_liste[1])
#> List of 1
  $ MonVecteur: int [1:10] 1 2 3 4 5 6 7 8 9 10
ma_liste[[1]]
   [1] 1 2 3 4 5 6 7 8 9 10
str(ma_liste[[1]])
#> int [1:10] 1 2 3 4 5 6 7 8 9 10
ma_liste$MonVecteur
   [1] 1 2 3 4 5 6 7 8 9 10
```

## Sélectionner les éléments d'une liste (2)

Il est possible qu'une liste contienne plusieurs éléments et les concepts de sélection sont les mêmes.

```
#> $niv3_1
#> [1] "Niveau 3"
#>
#> $niv3_2
#> $niv3_2$niv4_1
#> [1] "Niveau 4 - Élément 1"
#>
#> $niv3_2$niv4_2
#> [1] "Niveau 4 - Élément 2"

list_plusieurs_niveaux$niv1_2$niv2_2$niv3_2$niv4_2
```

#### **BRAVO!**

Vous êtes maintenant de jeunes padawans Rinessiens