

Rendu ARCHE

Name of the Student

Prise en main

1) Régardez la video : <https://rstudio.cloud/learn/primers/1.2>. Timming 20 Min.

Questions théoriques:

- Comment créer un vector sous R?

Votre réponse ici

```
# Mets ici un exemple
```

- Quelle est la fonction des crochets []? > Ecrire votre réponse ici

Exercice Pratique 1

On a demandé à 4 ménages le revenu des deux conjoints, et le nombre de personnes du ménage :

```
conjoint_1 <- c(1200, 1180, 1750, 2100, 1840)
conjoint_2 <- c(1450, 1870, 1690, 0, 2500)
nb_personnes <- c(4, 2, 3, 2, 5)
```

1. Calculer le revenu total de chaque ménage, puis diviser par le nombre de personnes pour obtenir le revenu par personne de chaque ménage.
2. • Calculer le revenu minimum et maximum parmi ceux du premier `conjoint_1`.
3. Calculer le revenu minimum et maximum, mais l'un des enquêtés n'a pas voulu répondre :

```
conjoint_1 <- c(1200, 1180, 1750, NA, 1840)
```

Exercice Pratique 2

On a relevé les notes en maths, anglais et sport d'une classe de 6 élèves et on a stocké ces données dans trois vecteurs :

```
maths <- c(12, 16, 8, 18, 6, 10)
anglais <- c(14, 9, 13, 15, 17, 11)
sport <- c(18, 11, 14, 10, 8, 12)
```

- Calculer la moyenne des élèves de la classe en anglais.
- Calculer la moyenne générale de chaque élève.
- Essayez de comprendre le résultat des deux fonctions suivantes (vous pouvez vous aider de la page d'aide de ces fonctions) :

```
pmin(maths, anglais, sport)
```

```
## [1] 12 9 8 10 6 10
```

```
pmax(maths, anglais, sport)
```

```
## [1] 18 16 14 18 17 12
```

Introduction au Tableau de Données

Régardez les tutoriels :

1. <https://rstudio.cloud/learn/primers/2.3>. Timming 20 Min.
2. <https://rstudio.cloud/learn/primers/2.3>

Questions théoriques

1. Qu'est-ce qu'un **tibble** par rapport au format **data.frame** et quel est la principale avantage de ce format ?

Votre réponse

2. Expliquez a quoi servent les fonctions suivantes (avec vos mots):

- **select()** and **filter()**: _____
- **arrange()**: _____
- **%>%**: _____
- **mutate()**, **group_by()**, and **summarise()**: _____

3. Quels sont les **helpers** de la fonction **select()**?

Votre réponse

4. Expliquez à quoi ça sert l'opérateur **>=**?

Votre réponse

5. Quels sont les quatre erreurs les plus courants avec la fonction **filter**?

Votre réponse

6. Expliquez l'erreur dans le code **filter(babynames, 10 < n < 20)**?

Votre réponse

Si je veux filtrer faire, comment il faut le faire?

```
# babynames %>% filter( _____ )
```

7. Expliquez l'interet de l'opérateur **Pipe %>%**?

Votre réponse

```
# Un exemple
```

Exercice Pratique 1

Vous avez ces donnees:

```
# Lire les donnees
mydata <- read_csv2("data/taille.csv")
```

Calculez:

- Donner le nombre total de réponse par sexe.

- Calculez la taille max, min et moyen par sexe.
- Filtrez les individus avec un Doctorat.
- Ajouter une nouvelle variable avec la notation “Good” pour grade A et “Neutral” grade B et “Bad” pour le grade C.

Introduction à la visualisation des données avec ggplot2

Régardez le tutoriel:

1. <https://rstudio.cloud/learn/primers/1.1>. Timming 20 Min.

- Quels sont les composants qu’un grapique a besoin en ggplot?

1. Votre réponse
2. Votre réponse
3. Votre réponse

- Qu’est ce que c’est un geom_ en ggplot ?

Votre réponse

- Faites un nuage de points entre les données cty vs hwy du jeu de données mpg?

```
# modifiez le code ci-dessous pour réaliser le nuage de points
# ggplot(data = <DATA>) +
#   aes(x = <X>, y = <Y>)
#   <GEOM_FUNCTION>()
```

- Dans le jeu de données mpg, graphiquez un nuage de points cty vs hwy, en prenant le facteur drv comme un variable de differentiation par la couleur.

```
#ggplot(data = mpg) +
#   aes(x = ____, y = _____, __ = _____ ) +
#   geom_point()
```

- Dans le jeu de données babynames, Grafiquez la popularité de prenom James et Emma.

```
#babynames %>% ..
```

- Peut-on grafiquer la curve de total des personnes différencié par sex depuis 1980?

```
# Code ici
```

Challenge Gapminder

```
# Code pour Lire la base de donnees 'Gapminder'
gapminder <- read_csv2("data/gapminder.csv")
```

```
## i Using ',' as decimal and '.' as grouping mark. Use `read_delim()` for more control.
```

```
##
```

```
## -- Column specification -----
```

```
## cols(
```

```
##   country = col_character(),
```

```
##   continent = col_character(),
```

```
##   year = col_double(),
```

```
##   lifeExp = col_double(),
```

```
##   pop = col_double(),
```

```
##   gdpPercap = col_double()
```

```
## )
```

```
#table  
gapminder
```

```
## # A tibble: 1,704 x 6
```

```
##   country      continent  year lifeExp      pop gdpPercap  
##   <chr>         <chr>    <dbl>  <dbl>    <dbl>    <dbl>  
## 1 Afghanistan Asia      1952   28.8  8425333    779.  
## 2 Afghanistan Asia      1957   30.3  9240934    821.  
## 3 Afghanistan Asia      1962   32.0 10267083    853.  
## 4 Afghanistan Asia      1967   34.0 11537966    836.  
## 5 Afghanistan Asia      1972   36.1 13079460    740.  
## 6 Afghanistan Asia      1977   38.4 14880372    786.  
## 7 Afghanistan Asia      1982   39.9 12881816    978.  
## 8 Afghanistan Asia      1987   40.8 13867957    852.  
## 9 Afghanistan Asia      1992   41.7 16317921    649.  
## 10 Afghanistan Asia      1997   41.8 22227415    635.  
## # ... with 1,694 more rows
```

Choisissez au moins **Deux** exercices ci-dessous en présentant un tableau ou bien un graphique

- dplyr devrait être votre principal outil de manipulation des données
- ggplot2 doit être votre principal outil de visualisation

Tasks menu

- Grafiquez l'espérance de vie moyenne sur l'ensemble des années pour la France ?
- Obtenir le maximum et le minimum du PIB par habitant pour tous les continents. (Fait un tableau)
- Regardez la répartition du PIB par habitant entre les pays au sein des continents.
- Comment l'espérance de vie varie-t-elle entre les différents continents ? (En utilisant ggplot)
- Proposez une analyse!