

Rendu ARCHE

Name of the Student

Prise en main

- Comment créer un vector sous R?

Votre réponse ici

```
# Mets ici un exemple
```

1. Quelle est la fonction des crochets []?

Ecrire votre réponse ici

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

```
conjoint1 <- c(1200, 1180, 1750, 2100)
conjoint2 <- c(1450, 1870, 1690, 0)
nb_personnes <- c(4, 2, 3, 2)
```

- Calculer le revenu total de chaque ménage, puis diviser par le nombre de personnes pour obtenir le revenu par personne de chaque ménage.
- Calculer le revenu minimum et maximum parmi ceux du premier conjoint.

```
conjoint1 <- c(1200, 1180, 1750, 2100)
```

1. Exercice 5

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.

ggplot

- 1 Quels sont les trois fonction essentielles pour créer un graphique en ggplot2?

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

1. Qu'est ce que c'est un geom_ en ggplot ?

Votre réponse

1. Quel geom_ utiliseriez-vous pour représenter un nuage de points ? Une boîte à moustaches ?

Votre réponse

```
# Une exemple de boîte à moustache donnees `mpg`
```

1. Quels sont les composant éléments qu'un graphics a besoin en ggplot?

```
# Votre réponse
```

1. Faites un nuage de points de `cty` vs `hwy` de jeu de donnees `mpg`?

```
# ggplot(data = <DATA>) +  
#   aes(x = <X>, y = <Y>)  
#   <GEOM_FUNCTION>()
```

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

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

Tidyverse

Tibble

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

Votre réponse

1. Expliquez les fonctions suivantes:

- `select()` and `filter()`: _____
- `arrange()`: _____
- `%>%`: _____
- `mutate()`, `group_by()`, and `summarize()`: _____

```
# Votre réponse
```

1. Quels sont les `helpers` de la fonctionne `select()`?

Votre réponse

1 Pouvez-vous écrire le code pour filtrer le prénom 'Zack', et savoir combien de fois dans il a été donné le jeu de donnees `babynames`?

Votre réponse

```
babynames %>% filter(name == 'James') %>% select(year, n)
```

```
## # A tibble: 276 x 2  
##   year      n  
##   <dbl> <int>  
## 1 1880     22  
## 2 1880    5927  
## 3 1881     24  
## 4 1881    5441  
## 5 1882     18  
## 6 1882    5892  
## 7 1883     25  
## 8 1883    5223
```

```
## 9 1884 33
## 10 1884 5693
## # ... with 266 more rows
```

1. Expliquez à quoi ça sert l'opérateur `>=`?

Votre réponse

1. Quels sont les quatre erreurs les plus courants avec la fonction `filter`?

Votre réponse

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

Votre réponse

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

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

1. Expliquez l'intérêt de l'opérateur **Pipe** `%>%`?

Votre réponse

```
# Un exemple
```

1. Quels sont les prénoms les plus populaires (masculin et féminin) l'année 2000 selon les données `babynames`?

```
# Ici votre code
```

```
babynames %>%
  filter(year == 2000, sex == 'F') %>%
  slice(desc(n))
```

```
## # A tibble: 16,751 x 5
##   year sex name      n      prop
##   <dbl> <chr> <chr>   <int>   <dbl>
## 1 2000 F Emily    25953 0.0130
## 2 2000 F Hannah   23080 0.0116
## 3 2000 F Madison  19967 0.0100
## 4 2000 F Ashley   17997 0.00902
## 5 2000 F Jordyn    1730 0.000867
## 6 2000 F Jazmine   1353 0.000678
## 7 2000 F Haylee    1261 0.000632
## 8 2000 F Miriam    1152 0.000577
## 9 2000 F Josephine 1114 0.000558
## 10 2000 F Deanna   1051 0.000527
## # ... with 16,741 more rows
```

1. Grafiquez la popularité de prénom `James`.

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

1. Listes les types de *fonctions de résumés* courantes dans statistique descriptive?

- 1.
- 2.

1. Quel est l'objectif de ce partie du code?. Expliquez

```
babynames %>%
  group_by(year, sex) %>%
  summarise(total = sum(n))
```

Votre réponse

1. Peut-on graphiquer la curve de total des personnes différencié par sexe depuis 1880?

Code ici

Challenge Gapminder

- Inspiré par Hans Rosling

Regardez la documentation de `gapminder` data package

```
# Code pour Lire la base de données '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 <- knitr::kable(head(gapminder, 10)) # knitr::kable is a function to plot beautiful tables in pd.
```

```
#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** tâches ci-dessous en présentant un tableau ou bien une graphique

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

Tasks menu

- 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!