

Rendu ARCHE

Nicolas SAILÉ

On R

1. Comment creer un vecteur sous R? Faite un exemple.

```
# Avec la fonction c(), par exemple :  
# vecteur = c(1, 2, 3)
```

2. Quelle est la fonction des crochets []?

```
# Accéder à l'un ou plusieurs des composants d'un vecteur.
```

3. Quelle est la fonction pour charger un package sous R?

```
# install.packages() pour installer puis library() pour charger
```

Questions pour ggplot

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

```
# "geom" est un objet géométrique, autrement dit il définit un type de représentation  
# graphique (nuage de points par exemple)
```

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

```
# geom_points() pour le nuage et geom_plotbox() pour la boite à moustache
```

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

```
# Il faut un set de données (data), un aesthetic (aes) définissant quelles données on veut  
# représenter (en x, en y, en couleurs, en taille...) et un objet géométrique (geom) pour  
# définir sous quelle forme nous voulons représenter ces données.
```

4. Qu'est-ce qu'un package ?

```
# C'est un ensemble de fonctions et données que l'on peut importer, qui ne sont  
# pour la plupart pas intégrées de base dans R pour économiser de la place.
```

Tidyverse

Tibble

1. Qu'est-ce qu'un tibble et quel est le principal avantage de ce format ?

```
# Les tibble permettent d'afficher des grandes quantités de données plus lisiblement  
# que de simples data frames grâce au package tibble
```

Dplyr

1. Expliquez avec vos mots la fonctionnalité des fonctions suivantes `select()`, `filter()` et `arrange()`?

```
# select() permet de n'afficher que certains paramètres (colonnes) voulus
# filter() permet d'imposer des conditions pour définir valeurs à afficher
# arrange() permet de trier les données selon leur ordre croissant/décroissant
# ou alphabétique (A->Z ou Z->A). On peut avoir plusieurs niveaux de tri
```

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

```
# Ce sont '-'; ':'; 'contains()'; 'ends_with()'; 'matches()'; 'num_range()';
# 'one_of()'; 'starts_with()'
```

- 1 Pouvez-vous écrire le code pour filtrer votre prénom dans le jeu de données `prenoms`?

```
# filter(prenoms, name == "Nicolas") //J'ai utilisé la version anglaise, j'utilise
# //donc "name" pour les prénoms
```

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

```
# L'élément précédent est supérieur ou égal à l'élément suivant
```

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

```
# = au lieu de ==
# string au lieu de "string"
# 10<a<20 au lieu de 10<a & a<20
# succession de OU logiques (|) au lieu de %in%
```

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

```
# cette condition doit être écrite en deux conditions distinctes (réponse précédente)
# On peut aussi vérifier que la valeur de n est %in% 10:20
```

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

```
# Il permet de définir sur une ligne le premier argument de la fonction suivante
```

1. Quel est le prénom le plus populaire pour masculin et féminin en 2000 dans le package `prenoms`?

```
# Dans la version anglaise "babynames", le plus populaire est
```

1. Grafiquez la popularité de votre prénom.

```
#babynames %>%
#   filter(name == "Nicolas", sex == "M") %>%
#   select(year,prop) %>%
#   ggplot() +
#     aes(year, prop) +
#     geom_line()
```

```
#Le résultat est donné sur la figure 1
```

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

```
# Nous avons vu mean et sd pour générer des nombres normalement
# distribués. Il existe de nombreuses fonctions statistiques descriptives comme par
# exemple min(), max(), median(), var(), nth() etc...
#
# Je n'avais pas encore lu la partie "Derive information" : les fonctions de résumés
```

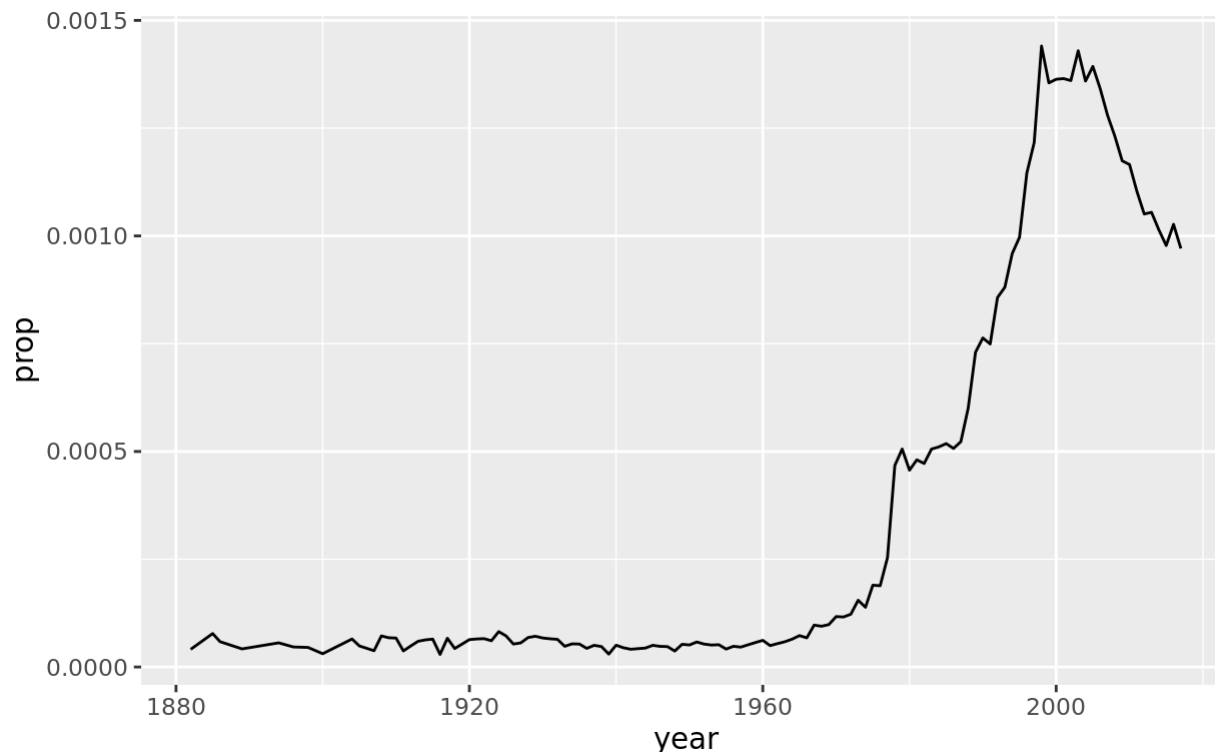


Figure 1: Graphique de l'évolution de la popularité du prénom Nicolas dans les données “babynames”

*# sont mean(x), median(x), quantile(x, 0.25), min(x), max(x), sd(x), var(x), IQR(x),
mad(x), first(x), nth(x, 2), last(x), n_distinct(x) et éventuellement n().
On peut également y ajouter des conditions logiques.*

1. Dans le **Challenge “prénoms distincts”**, A votre avis, ces chiffres seront-ils différents ? Pourquoi ?

*# Oui les chiffres sont différents car un même prénom peut apparaître plusieurs fois
dans différentes années et compter comme une ligne à chaque fois.*

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

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

*# Ce code donne le nombre total de naissance par genre et par année sous la forme
d'un tibble à 3 colonnes*

Tidy data

1. Décrire les trois critères d'un jeu de données en mode tidy data:

*# Un jeu de données est en tidy data si
- Il y a une variable par colonne
- Il y a une observation (mesure) par ligne
- Chaque valeur est dans sa propre cellule*

1. Quel est l'intérêt des fonctions `pivot-longer()` et `pivot-wider()` ?

```
# pivot-longer() : passer d'un format large à un format long  
# pivot-wider() : passer d'un format long à un format large
```

1. Dans l'utilisation de `pivot_longer()` et `pivot_wider()`, expliquez la signification des guillemets.

```
# Elle permet d'indiquer si l'on veut parler de la valeur d'une variable ou du nom  
# de cette variable. Par exemple pour la colonne intitulée "1999".
```

1. Dans l'exercice pour calculer le ratio garçons/filles au cours du temps. Quel interprétation donnerez-vous à cette graphique

```
# J'avais supprimé cette partie à cause du problème d'export pdf mais si je me  
# souviens bien, la courbe de ratio était croissante ce qui signifie qu'il y a  
# de plus en plus de garçons que de filles.
```