

# Rapport d'étude

Papa Amadou NIANG

2024-03-08

## Présentation de la base de données

Dans une étude fictive, 1000 profils ont été simulés, reflétant une diversité de caractéristiques démographiques et socio-professionnelles. Les variables incluent le sexe, la taille, le poids, et l'âge, avec des distributions normales et uniformes. Les données s'étendent sur des aspects tels que le salaire, la profession, la ville de résidence, le niveau d'éducation, le statut familial, les habitudes de vie, le bien-être, le groupe sanguin, la vaccination contre le Covid-19, la distance domicile-travail, et le type de logement. Cette base de données imaginaire offre un terrain fertile pour des analyses statistiques multidimensionnelles.

```
# Définir le nombre d'observations
n <- 1000

# Générer des données aléatoires
data <- data.frame(
  Sexe = sample(c("Homme", "Femme"), n, replace = TRUE),
  Taille = rnorm(n, mean=170, sd=10),
  Poids = rnorm(n, mean=70, sd=15),
  Age = sample(18:100, n, replace = TRUE),
  Salaire = runif(n, min=20000, max=100000),
  Profession = sample(c("Enseignant", "Ingénieur", "Médecin",
    "Avocat", "Artiste"), n, replace = TRUE),
  Ville = sample(c("Dakar", "Thiès", "Kaolack", "Ziguinchor",
    "Touba"), n, replace = TRUE),
  Education = sample(c("Secondaire", "Bac", "Licence",
    "Master", "Doctorat"), n, replace = TRUE),
  Enfant = sample(0:5, n, replace = TRUE),
  Fumeur = sample(c("Oui", "Non"), n, replace = TRUE),
  Sportif = sample(c("Oui", "Non"), n, replace = TRUE),
  ScoreSatisfaction = runif(n, 0, 100),
  GroupeSanguin = sample(c("A", "B", "AB", "O"), n, replace = TRUE),
  VaccinéCovid = sample(c(TRUE, FALSE), n, replace = TRUE),
  DistanceTravail = rnorm(n, mean=10, sd=5),
  TypeLogement = sample(c("Maison", "Appartement",
    "Studio"), n, replace = TRUE)
)
summary(data)
```

##	Sexe	Taille	Poids	Age
##	Length:1000	Min. :137.1	Min. : 22.38	Min. : 18.00
##	Class :character	1st Qu.:163.0	1st Qu.: 59.18	1st Qu.: 40.00
##	Mode :character	Median :170.1	Median : 70.17	Median : 62.00

```
##           Mean    :169.9   Mean    : 69.76   Mean    : 60.42
##           3rd Qu.:176.8   3rd Qu.: 79.66   3rd Qu.: 81.00
##           Max.    :198.5   Max.    :119.86   Max.    :100.00
##   Salaire      Profession      Ville      Education
##   Min.    :20085   Length:1000   Length:1000   Length:1000
##   1st Qu.:41284   Class :character   Class :character   Class :character
##   Median :61272   Mode  :character   Mode  :character   Mode  :character
##   Mean    :61593
##   3rd Qu.:82315
##   Max.    :99997
##   Enfant      Fumeur      Sportif      ScoreSatisfaction
##   Min.    :0.000   Length:1000   Length:1000   Min.    : 0.00037
##   1st Qu.:1.000   Class :character   Class :character   1st Qu.:23.94529
##   Median :3.000   Mode  :character   Mode  :character   Median :47.94590
##   Mean    :2.506
##   3rd Qu.:4.000
##   Max.    :5.000
##   GroupeSanguin  VaccinéCovid  DistanceTravail  TypeLogement
##   Length:1000    Mode :logical   Min.    :-10.416   Length:1000
##   Class :character  FALSE:524   1st Qu.: 6.636   Class :character
##   Mode  :character  TRUE :476    Median : 10.107   Mode  :character
##                                     Mean    : 10.009
##                                     3rd Qu.: 13.423
##                                     Max.    : 27.173
```

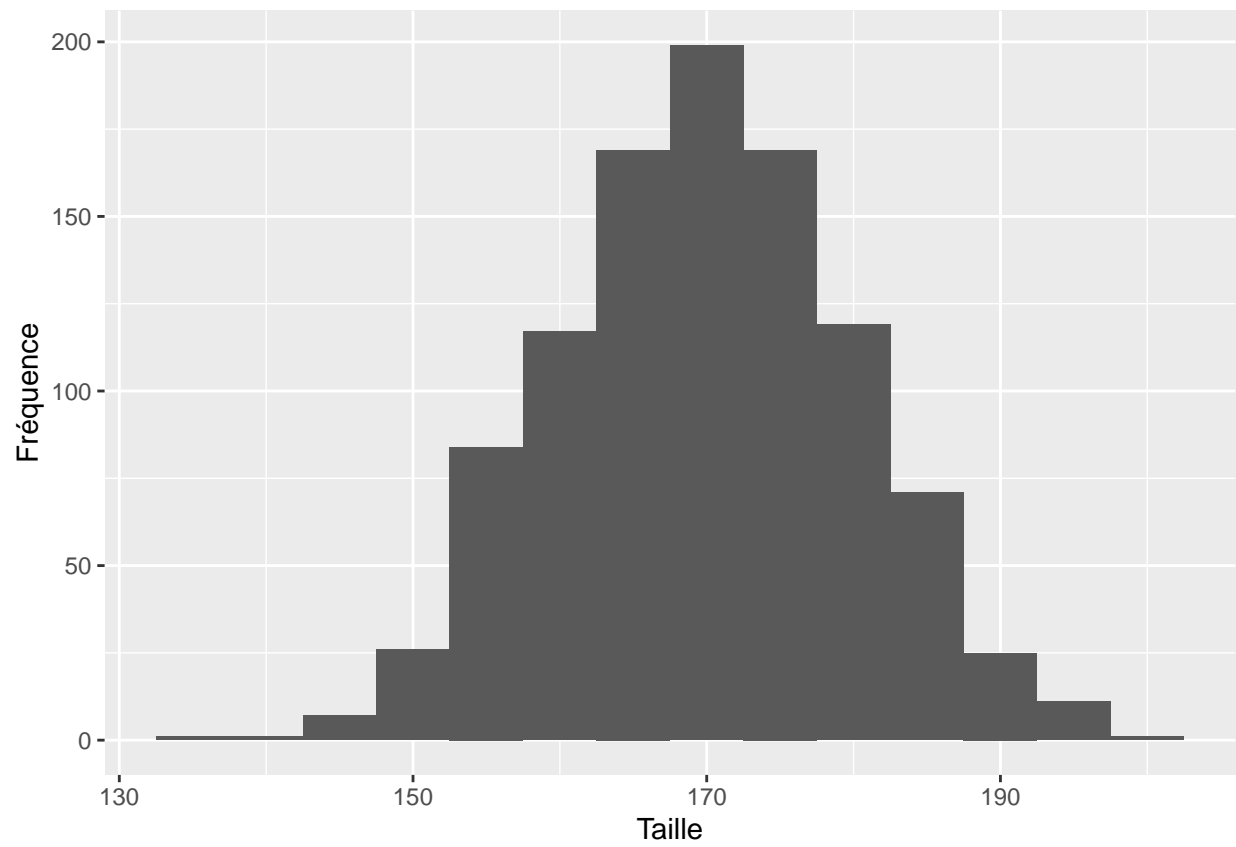
## Statistiques descriptives

Nous commençons par examiner les variables quantitatives de notre ensemble de données. La taille moyenne des individus est de 169.90378 cm, avec un écart-type de 9.9702797. Cela indique une distribution relativement homogène autour de la moyenne. En ce qui concerne le poids, la moyenne est de 69.7573457 kg, accompagnée d'un écart-type de 14.951409, reflétant une variabilité similaire à celle de la taille.

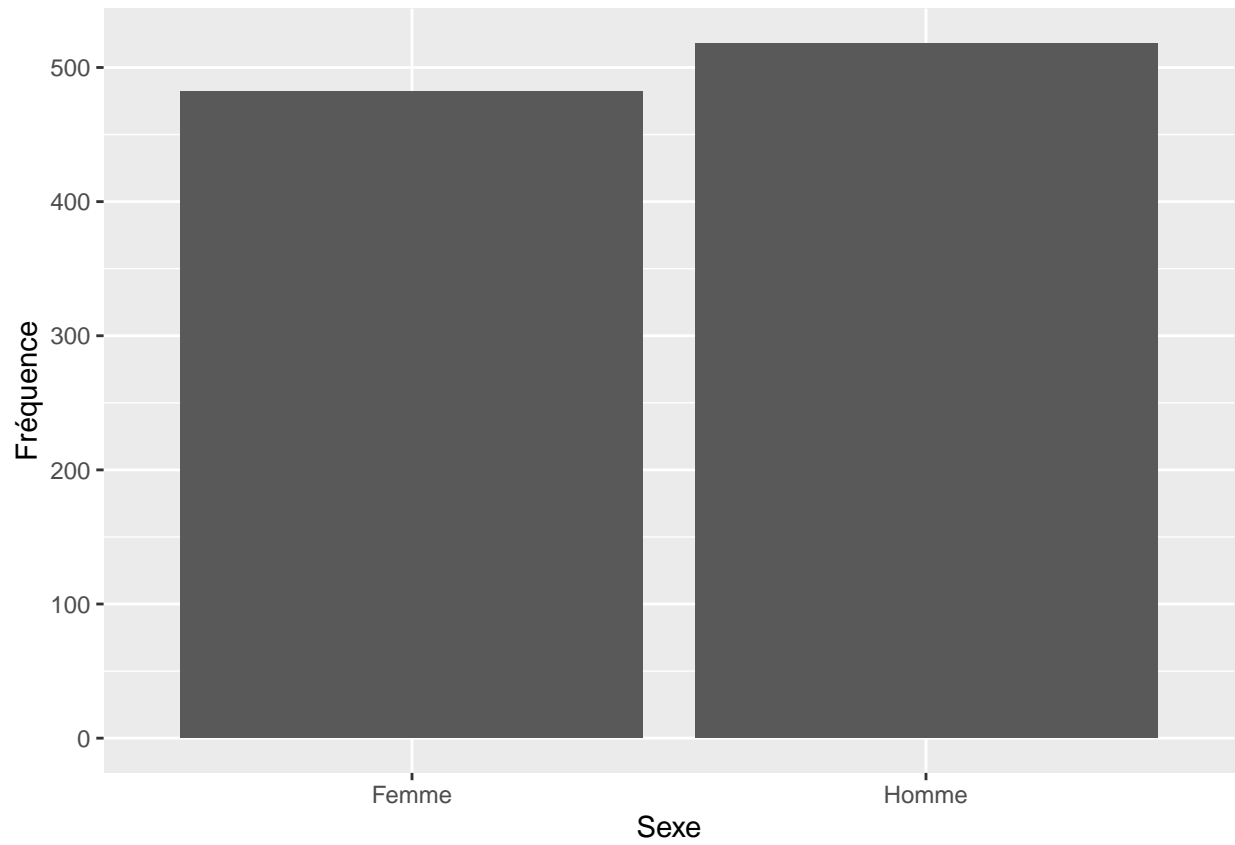
L'âge moyen des participants est de 60.424 ans, ce qui suggère une population d'adultes avec une présence notable de personnes âgées, comme le montre l'écart-type de 23.7825531 ans. Le salaire annuel moyen s'élève à  $6.1592804 \times 10^4$  euros, avec une dispersion mesurée par un écart-type de  $2.3374953 \times 10^4$  euros, indiquant une disparité économique au sein de la population.

## Visualisation des données

```
library(ggplot2)
ggplot(data = data, aes(x = Taille)) +
  geom_histogram(binwidth = 5) +
  xlab("Taille") +
  ylab("Fréquence")
```



```
ggplot(data = data, aes(x = Sexe)) +  
  geom_bar() +  
  xlab("Sexe") +  
  ylab("Fréquence")
```



```
ggplot(data = data, aes(x = "", y = ScoreSatisfaction)) +  
  geom_boxplot() +  
  ylab("Score de Satisfaction")
```

