Exercice 4

Tamsir Ndong

2024-03-08

## R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

n<-1000  
  
# Générer des données aléatoires  
data <- data.frame(  
 Sexe = sample(c("Homme", "Femme"), n, replace = TRUE),  
 Taille = rnorm(n, mean=170, sd=10), # Taille en cm  
 Poids = rnorm(n, mean=70, sd=15), # Poids en kg  
 Age = sample(18:100, n, replace = TRUE), # Age en années  
 Salaire = runif(n, min=20000, max=100000), # Salaire annuel en euros  
 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), # Nombre d'enfants  
 Fumeur = sample(c("Oui", "Non"), n, replace = TRUE),  
 Sportif = sample(c("Oui", "Non"), n, replace = TRUE),  
 ScoreSatisfaction = runif(n, 0, 100), # Score de satisfaction  
 GroupeSanguin = sample(c("A", "B", "AB", "O"), n, replace = TRUE), # Groupe sanguin  
 VaccinéCovid = sample(c(TRUE, FALSE), n, replace = TRUE), # Vacciné contre le Covid-19  
 DistanceTravail = rnorm(n, mean=10, sd=5), # Distance du travail en km  
 TypeLogement = sample(c("Maison", "Appartement", "Studio"), n, replace = TRUE) # Type de logement  
)  
  
  
summary(data)

## Sexe Taille Poids Age   
## Length:1000 Min. :141.1 Min. : 28.85 Min. : 18.00   
## Class :character 1st Qu.:163.8 1st Qu.: 60.07 1st Qu.: 37.00   
## Mode :character Median :170.6 Median : 69.98 Median : 59.50   
## Mean :170.5 Mean : 69.70 Mean : 58.59   
## 3rd Qu.:176.9 3rd Qu.: 79.88 3rd Qu.: 79.00   
## Max. :200.6 Max. :116.94 Max. :100.00   
## Salaire Profession Ville Education   
## Min. :20066 Length:1000 Length:1000 Length:1000   
## 1st Qu.:38821 Class :character Class :character Class :character   
## Median :58837 Mode :character Mode :character Mode :character   
## Mean :59142   
## 3rd Qu.:79676   
## Max. :99960   
## Enfant Fumeur Sportif ScoreSatisfaction  
## Min. :0.000 Length:1000 Length:1000 Min. : 0.1392   
## 1st Qu.:1.000 Class :character Class :character 1st Qu.:26.3799   
## Median :3.000 Mode :character Mode :character Median :50.0795   
## Mean :2.497 Mean :50.2662   
## 3rd Qu.:4.000 3rd Qu.:74.3930   
## Max. :5.000 Max. :99.8370   
## GroupeSanguin VaccinéCovid DistanceTravail TypeLogement   
## Length:1000 Mode :logical Min. :-5.741 Length:1000   
## Class :character FALSE:508 1st Qu.: 6.514 Class :character   
## Mode :character TRUE :492 Median : 9.832 Mode :character   
## Mean : 9.796   
## 3rd Qu.:13.130   
## Max. :27.715

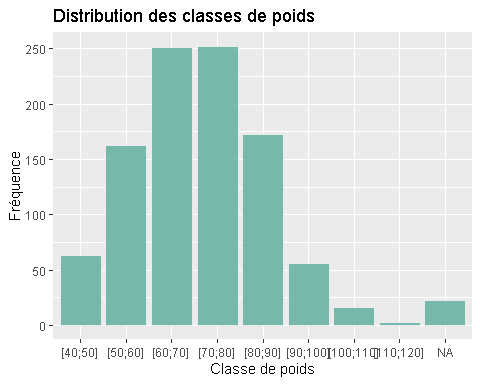
## Including Plots

# Définir les limites des classes pour l'âge, la taille et le poids  
limites\_age <- seq(0, 100, by = 10) # Classes d'âge de 10 ans  
limites\_taille <- seq(120, 200, by = 10) # Classes de taille de 10 cm  
limites\_poids <- seq(40, 120, by = 10) # Classes de poids de 10 kg  
  
# Définir les labels pour chaque classe  
labels\_age <- paste("[", head(limites\_age, -1), ";", tail(limites\_age, -1), "]", sep = "")  
labels\_taille <- paste("[", head(limites\_taille, -1), ";", tail(limites\_taille, -1), "]", sep = "")  
labels\_poids <- paste("[", head(limites\_poids, -1), ";", tail(limites\_poids, -1), "]", sep = "")  
  
# Créer des variables de classe pour l'âge, la taille et le poids  
data$ClasseAge <- cut(data$Age, breaks = limites\_age, include.lowest = TRUE, labels = labels\_age)  
data$ClasseTaille <- cut(data$Taille, breaks = limites\_taille, include.lowest = TRUE, labels = labels\_taille)  
data$ClassePoids <- cut(data$Poids, breaks = limites\_poids, include.lowest = TRUE, labels = labels\_poids)

You can also embed plots, for example:

## Warning: le package 'ggplot2' a été compilé avec la version R 4.3.3

## Warning in geom\_histogram(stat = "count", fill = "#69b3a2", color = "#e9ecef",  
## : Ignoring unknown parameters: `binwidth`, `bins`, and `pad`



Note that the echo = FALSE parameter was added to the code chunk to prevent printing of the R code that generated the plot.

# Calculer le tableau de contingence (classeAge et fumeur)  
tableau1 <- table(data$ClasseAge, data$Fumeur)  
  
print(tableau1)

##   
## Non Oui  
## [0;10] 0 0  
## [10;20] 21 24  
## [20;30] 71 48  
## [30;40] 60 60  
## [40;50] 53 58  
## [50;60] 57 56  
## [60;70] 60 79  
## [70;80] 47 80  
## [80;90] 59 50  
## [90;100] 64 53