

RÉPUBLIQUE DU SÉNÉGAL

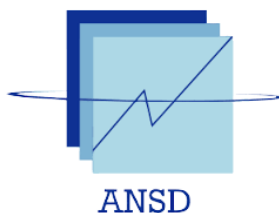


Un Peuple - Un But - Une Foi

Ministère de l'Économie, du Plan et de la Coopération



Agence Nationale de la Statistique et de la Démographie (ANSD)



École Nationale de la Statistique et de l'Analyse Économique Pierre Ndiaye (ENSAE)



PROJET STATISTIQUES SOUS R TP1 : Documentation en Rmarkdown

Rédigé par :

Mame Balla BOUSSO

Ameth FAYE

EDIMA Biyenda Hildégarde

Papa Amadou NIANG

Elèves ingénieurs statisticiens économistes

Sous la supervision de :

M. Aboubacar HEMA

ANALYSTE DE RECHERCHE CHEZ

IFPRI

Année scolaire : 2024/2025

I. Introduction et prise en main

Qu'est-ce que R Markdown ?

Définition :

R Markdown est un format de document dynamique qui permet d'intégrer du code R avec du texte explicatif dans un même fichier. Il est basé sur la syntaxe Markdown, une syntaxe légère pour le formatage de texte, et il utilise des outils comme `knitr` et `pandoc` pour générer des rapports, des présentations, et même des sites web.

Avantages de R Markdown :

- **Reproductibilité** : Générer des rapports automatiques et reproductibles.
- **Formats variés** : Peut être converti en HTML, PDF, Word, etc.
- **Intégration de code** : Permet d'exécuter du code R directement dans le document.
- **Automatisation** : Idéal pour les rapports récurrents.
- **Clarté** : Séparation du code et du texte explicatif.

Pourquoi utiliser R Markdown ?

R Markdown est particulièrement utile dans le domaine des statistiques et de l'analyse de données, car il permet de combiner la puissance de R pour le traitement des données avec une présentation claire et structurée des résultats. Que vous souhaitiez créer des rapports académiques, des analyses de données interactives ou des présentations professionnelles, R Markdown offre la flexibilité nécessaire pour répondre à vos besoins.

II. Les bases de R Markdown

1. Syntaxe Markdown

R Markdown repose sur Markdown, qui permet d'écrire du texte en utilisant une syntaxe simple pour la mise en forme.

a) Mise en forme inline (italique, gras, liens, etc.)

Voici quelques exemples de mise en forme :

- **Italique** :

```
cat("*Texte en italique* ou _Texte en italique_")
```

```
## *Texte en italique* ou _Texte en italique_
```

Texte en italique ou *Texte en italique*

- **Gras** :

```
cat("***Texte en gras** ou __Texte en gras__")
```

```
## **Texte en gras** ou __Texte en gras__
```

Texte en gras ou **Texte en gras**

- **Lien :**

```
cat("[Nom du lien](https://github.com/Abson-dev/Projet-statistique-sous-R)")
```

```
## [Nom du lien](https://github.com/Abson-dev/Projet-statistique-sous-R)
```

Nom du lien

b) Éléments de niveau bloc (titres, listes, citations...)

- **Titres :**

```
cat("# Titre de niveau 1\n## Titre de niveau 2\n### Titre de niveau 3")
```

Titre de niveau 1

Titre de niveau 2

Titre de niveau 3

- **Listes à puces :**

```
cat("- Élément 1\n- Élément 2\n  - Sous-élément 2.1")
```

- Élément 1
- Élément 2
 - Sous-élément 2.1
- Élément 1
- Élément 2
 - Sous-élément 2.1

- **Listes numérotées :**

```
cat("1. Premier point\n2. Deuxième point")
```

1. Premier point
2. Deuxième point
3. Premier point
4. Deuxième point

- **Citations :**

```
cat("> Ceci est une citation")
```

Ceci est une citation

Ceci est une citation

c) Expressions mathématiques

R Markdown permet d'intégrer des équations LaTeX pour les formules mathématiques :

-Binôme de Newton

```
cat("Le théorème du binôme de Newton s'écrit comme suit : \n\n")
```

Le théorème du binôme de Newton s'écrit comme suit :

```
cat("$$ (a + b)^n = \sum_{k=0}^n \binom{n}{k} a^{n-k} b^k $$\n\n")
```

$$(a + b)^n = \sum_{k=0}^n \binom{n}{k} a^{n-k} b^k$$

-Fonction continue

```
cat("Une fonction \n( f \n) est continue en un point \n( x_0 \n) si : \n\n")
```

Une fonction f est continue en un point x_0 si :

```
cat("$$ \lim_{x \to x_0} f(x) = f(x_0) $$\n\n")
```

$$\lim_{x \rightarrow x_0} f(x) = f(x_0)$$

-Topologie

```
cat("Dans un espace métrique \n( (E, d) \n), la boule ouverte de centre \n( a \n) e
```

Dans un espace métrique (E, d) , la boule ouverte de centre a et de rayon $r > 0$ est définie par :

```
cat("$$ B(a, r) = \{ x \in E \mid d(x, a) < r \} $$\n\n")
```

$$B(a, r) = \{x \in E \mid d(x, a) < r\}$$

-Equations différentielles

```
cat("Une équation différentielle d'ordre \n( n \n) s'exprime sous la forme général
```

Une équation différentielle d'ordre n s'exprime sous la forme générale :

```
cat("$$ F \left( x, y, \frac{dy}{dx}, \frac{d^2y}{dx^2}, \dots, \frac{d^ny}{dx^n} \right) = 0
```

$$F \left(x, y, \frac{dy}{dx}, \frac{d^2y}{dx^2}, \dots, \frac{d^ny}{dx^n} \right) = 0$$

-Théorie des sondages

```
cat("Dans la théorie des sondages, l'estimateur de la moyenne de la population est
```

Dans la théorie des sondages, l'estimateur de la moyenne de la population est donné par :

```
cat("$$ \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i $$\n\n")
```

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

-Algèbre linéaire

`cat("En algèbre linéaire, une matrice $\left(A \right)$ de dimension $\left(m \times n \right)$ es`

En algèbre linéaire, une matrice A de dimension $m \times n$ est représentée sous la forme suivante :

```
cat("$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \n")
```

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

-Probabilités et statistiques

`cat("L'espérance mathématique d'une variable aléatoire (X) discrète est donn`

L'espérance mathématique d'une variable aléatoire X discrète est donnée par :

```
cat("$\mathbb{E}[X] = \sum_{i=1}^n x_i P(x_i) \n")
```

$$\mathbb{E}[X] = \sum_{i=1}^n x_i P(x_i)$$

`cat("La variance est définie comme suit : \n")`

La variance est définie comme suit :

```
cat("$\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 \n")
```

$$\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$$

-Séries de Fourier

`cat("La décomposition en série de Fourier d'une fonction périodique $(f(x))$ e`

La décomposition en série de Fourier d'une fonction périodique $f(x)$ est donnée par :

```
cat("$f(x) = a_0 + \sum_{n=1}^{\infty} \left( a_n \cos(nx) + b_n \sin(nx) \right) \n")
```

$$f(x) = a_0 + \sum_{n=1}^{\infty} (a_n \cos(nx) + b_n \sin(nx))$$

`cat("Où les coefficients sont définis par : \n")`

Où les coefficients sont définis par :

```
cat("$a_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos(nx) dx \n")
```

$$a_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos(nx) dx$$

```
cat("$$ b_n = \\frac{1}{\\pi} \\int_{-\\pi}^{\\pi} f(x) \\sin(nx) \\, dx $$\\n\\n")
```

$$b_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin(nx) dx$$

-Séries de Taylor

```
cat("Le développement en série de Taylor d'une fonction \\( f(x) \\) autour d'un p
```

Le développement en série de Taylor d'une fonction $f(x)$ autour d'un point a est donné par :

```
cat("$$ f(x) = f(a) + f'(a)(x-a) + \\frac{f''(a)}{2!}(x-a)^2 + \\dots $$\\n\\n")
```

$$f(x) = f(a) + f'(a)(x - a) + \frac{f''(a)}{2!}(x - a)^2 + \dots$$

2. Table des matières et options globales

Vous pouvez configurer une table des matières et d'autres options globales en modifiant les métadonnées YAML et le chunk `setup` au début du document.

Ajouter une table des matières et numéroté les sections

Les options pour la table des matières et la numérotation des sections sont déjà incluses dans les métadonnées YAML :

```
output:
  pdf_document:
    latex_engine: xelatex
    keep_tex: true
    toc: true
    number_sections: true
```

III. Formats de sortie courants

La version originale de Markdown a été inventée principalement pour écrire plus facilement du contenu HTML. Par exemple, vous pouvez écrire une puce avec `- text` au lieu du code HTML détaillé `text`, ou une citation avec `> text` au lieu de `<blockquote>text</blockquote>`.

La syntaxe de Markdown a été considérablement étendue par Pandoc. De plus, Pandoc permet de convertir un document Markdown en une grande variété de formats de sortie. Dans cette partie, nous présenterons les fonctionnalités de divers formats de sortie de document.

1. Document HTML

Comme nous venons de le mentionner, Markdown a été conçu à l'origine pour la sortie HTML, il n'est donc pas surprenant que le format HTML possède les fonctionnalités les plus riches parmi tous les formats de sortie.

Pour créer un document HTML à partir de R Markdown, vous spécifiez le `html_document` format de sortie dans les métadonnées YAML de votre document :

```

---
title: "Titre du document"
author: Auteur
date: La date
output: html_document
---
```

1.1 Table des matières

Vous pouvez ajouter une table des matières (TOC) à l'aide de l' `toc` option et spécifier la profondeur des en-têtes à appliquer à l'aide de l' `toc_depth` option. Par exemple :

```

---
title: "Titre du document"
output:
  html_document:
    toc: true
    toc_depth: 2
---
```

Si la profondeur de la table des matières n'est pas explicitement spécifiée, la valeur par défaut est 3 (ce qui signifie que tous les en-têtes de niveau 1, 2 et 3 seront inclus dans la table des matières).

1.2 Numérotation des sections

Vous pouvez ajouter une numérotation de section aux en-têtes à l'aide de l' `number_sections` option :

```

---
title: "Titre du document"
output:
  html_document:
    toc: true
    number_sections: true
---
```

Notez que si vous choisissez d'utiliser cette `number_sections` option, vous souhaiterez probablement également utiliser # des en-têtes (H1) dans votre document, car ## les en-têtes (H2) incluront un point décimal, car sans en-têtes H1, vos en-têtes H2 seront numérotés avec 0.1, 0.2, etc.

1.3 Apparence et style

Il existe plusieurs options qui contrôlent l'apparence des documents HTML :

- **theme** spécifie le thème Bootstrap à utiliser pour la page (les thèmes sont tirés de la bibliothèque de thèmes Bootswatch). Les thèmes valides incluent default, bootstrap, cerulean, cosmo, darkly, flatly, journal, lumen, paper, readable, sandstone, simplex, spacelab, united et yeti. Ne passez `null` aucun thème (dans ce cas, vous pouvez utiliser le `css` paramètre pour ajouter vos propres styles).
- **highlight** spécifie le style de mise en surbrillance de la syntaxe. Les styles pris en charge incluent default, tango, pygments, kate, monochrome, espresso, zenburn, haddock, breezedark, et textmate. Passez `null` pour empêcher la mise en surbrillance de la syntaxe.
- **smart** indique si une sortie typographiquement correcte doit être produite, en convertissant les guillemets droits en guillemets bouclés, — en tirets cadratins, – en tirets demi-cadratins et ... en points de suspension. Notez que cette `smart` option est activée par défaut.

Par exemple :

```
---
title: "Titre du document"
output:
  html_document:
    theme: united
    highlight: tango
---
```

1.4 CSS personnalisé

Vous pouvez ajouter votre propre CSS à un document HTML en utilisant l' `css` option :

```
---
title: "Titre du document"
output:
  html_document:
    css: styles.css
---
```

2. Document PDF

Pour créer un document PDF à partir de R Markdown, vous spécifiez le `pdf_document` format de sortie dans les métadonnées YAML :

```
---
title: "Titre du document"
author: Auteur
date: La date
output: pdf_document
---
```

Dans les documents R Markdown qui génèrent une sortie PDF, vous pouvez utiliser du LaTeX brut et même définir des macros LaTeX. Consultez la documentation de Pandoc sur l'extension `raw_tex` pour plus de détails.

Notez que la sortie PDF (y compris les diapositives Beamer) nécessite une installation de LaTeX.

2.1 Table des matières

Vous pouvez ajouter une table des matières à l'aide de l' `toc` option et spécifier la profondeur des en-têtes à appliquer à l'aide de l' `toc_depth` option. Par exemple :

```
---
title: "Titre du document"
output:
  pdf_document:
    toc: true
    toc_depth: 2
---
```

Si la profondeur de la table des matières n'est pas explicitement spécifiée, elle est par défaut de 2 (ce qui signifie que tous les en-têtes de niveau 1 et 2 seront inclus dans la table des matières), tandis qu'elle est par défaut de 3 dans `html_document`.

Vous pouvez ajouter une numérotation de section aux en-têtes à l'aide de l' `number_sections` option :


```

---
title: "Titre du document"
output:
  pdf_document:
    toc: true
    number_sections: true
---

```

Si vous êtes familier avec LaTeX, `number_sections: true` signifie `\section{}`, et `number_sections: false` signifie `\section*{}` pour les sections dans LaTeX (cela s'applique également à d'autres niveaux de « sections » tels que `\chapter{}`, et `\subsection{}`).

2.2 Option LaTeX

De nombreux aspects du modèle LaTeX utilisé pour créer des documents PDF peuvent être personnalisés à l'aide de métadonnées YAML de niveau supérieur (notez que ces options n'apparaissent pas sous la section `output`, mais apparaissent plutôt au niveau supérieur avec `title`, `author`, etc.). Par exemple :

```

---
title: "Titre du document"
output: pdf_document
fontsize: 11pt
geometry: margin=1in
---

```

3. Document Word

Pour créer un document Word à partir de R Markdown, vous spécifiez le `word_document` format de sortie dans les métadonnées YAML de votre document :

```

---
title: "Titre du document"
author: Auteur
date: La date
output: word_document
---

```

La fonctionnalité la plus remarquable des documents Word est le modèle Word, également connu sous le nom de « document de référence de style ». Vous pouvez spécifier un document à utiliser comme référence de style lors de la production d'un `*.docx` fichier (un document Word). Cela vous permettra de personnaliser des éléments tels que les marges et d'autres caractéristiques de mise en forme. Pour de meilleurs résultats, le document de référence doit être une version modifiée d'un `.docx` fichier produit à l'aide de `rmarkdown` ou de `Pandoc`. Le chemin d'accès d'un tel document peut être passé à l'argument `reference_docx` du `word_document` format. Passez `"default"` pour utiliser les styles par défaut. Par exemple :

```

---
title: "Titre du document"
author: Auteur
date: La date
output:
  word_document:
    reference_docx: "mon_modele.docx"
---

```

IV. Création et manipulation de bases de données fictives

Pour illustrer les concepts de R Markdown, nous allons créer une base de données fictive qui servira de base pour nos tableaux, graphiques et équations statistiques. Cette base de données représentera un échantillon de données démographiques et économiques.

1. Création de la base de données

Nous allons créer un dataframe nommé `df_fictif` contenant les variables suivantes :

- **ID** : Identifiant unique de l'individu.
- **Age** : Âge de l'individu.
- **Sexe** : Sexe de l'individu (Homme/Femme).
- **Revenu** : Revenu mensuel.
- **Education** : Niveau d'éducation (Primaire, Secondaire, Universitaire).
- **Ville** : Ville de résidence (Ville A, Ville B, Ville C).
- **Nombre_Enfants** : Nombre d'enfants.

```
# Chargement des packages nécessaires
library(dplyr)
library(ggplot2)
library(knitr)
library(kableExtra)
library(tidyr)
library(xtable)

set.seed(42) # Pour la reproductibilité

# Création du dataframe
df_fictif <- data.frame(
  ID = 1:500,
  Age = sample(18:65, 500, replace = TRUE),
  Sexe = sample(c("Homme", "Femme"), 500, replace = TRUE, prob = c(0.49, 0.51)),
  Revenu = round(rnorm(500, mean = 3000, sd = 800), 2),
  Education = sample(c("Primaire", "Secondaire", "Universitaire"), 500, replace = TRUE),
  Ville = sample(c("Ville A", "Ville B", "Ville C"), 500, replace = TRUE, prob = c(0.3, 0.25, 0.2, 0.15, 0.1)),
  Nombre_Enfants = sample(0:5, 500, replace = TRUE, prob = c(0.3, 0.25, 0.2, 0.15, 0.1))
)

# Visualisation des premières lignes
head(df_fictif)
```

	ID	Age	Sexe	Revenu	Education	Ville	Nombre_Enfants
## 1	1	54	Homme	2443.58	Secondaire	Ville A	3
## 2	2	18	Homme	2719.92	Secondaire	Ville B	4
## 3	3	42	Femme	2848.79	Primaire	Ville A	0
## 4	4	27	Homme	3390.21	Primaire	Ville A	0
## 5	5	53	Femme	3890.67	Secondaire	Ville A	0
## 6	6	35	Femme	1219.94	Secondaire	Ville A	0

2. Aperçu de la base de données

Table 1: Aperçu de la base de données fictive

ID	Age	Sexe	Revenu	Education	Ville	Nombre_Enfants
1	54	Homme	2443.58	Secondaire	Ville A	3
2	18	Homme	2719.92	Secondaire	Ville B	4
3	42	Femme	2848.79	Primaire	Ville A	0
4	27	Homme	3390.21	Primaire	Ville A	0
5	53	Femme	3890.67	Secondaire	Ville A	0
6	35	Femme	1219.94	Secondaire	Ville A	0

3. Résumé statistique

Obtenons un résumé statistique des variables numériques de notre base de données.

```
summary(df_fictif)
```

```
##           ID           Age           Sexe           Revenu
##  Min.      : 1.0      Min.    :18.00      Length:500      Min.      : 631.3
## 1st Qu.:125.8      1st Qu.:31.00      Class :character  1st Qu.:2447.5
## Median :250.5      Median :43.00      Mode  :character  Median :3050.3
## Mean   :250.5      Mean   :42.39                      Mean   :3043.5
## 3rd Qu.:375.2      3rd Qu.:54.00                      3rd Qu.:3576.4
## Max.    :500.0      Max.    :65.00                      Max.    :5557.7
## Education           Ville           Nombre_Enfants
## Length:500          Length:500      Min.      :0.000
## Class :character     Class :character  1st Qu.:0.000
## Mode  :character     Mode  :character  Median :1.000
##                               Mean   :1.634
##                               3rd Qu.:3.000
##                               Max.    :5.000
```

V. Travailler avec des tableaux en R Mark-down

Les tableaux sont essentiels pour présenter des données de manière claire et structurée. R Markdown offre plusieurs moyens de créer et de styliser des tableaux.

1. Tableaux simples avec `knitr::kable`

La fonction `kable` du package `knitr` permet de créer des tableaux simples et esthétiques.

```
kable(head(df_fictif, 10), caption = "Tableau simple des 10 premières entrées") %>%
  kable_styling(full_width = FALSE, position = "center")
```

Table 2: Tableau simple des 10 premières entrées

ID	Age	Sexe	Revenu	Education	Ville	Nombre_Enfants
1	54	Homme	2443.58	Secondaire	Ville A	3
2	18	Homme	2719.92	Secondaire	Ville B	4
3	42	Femme	2848.79	Primaire	Ville A	0
4	27	Homme	3390.21	Primaire	Ville A	0
5	53	Femme	3890.67	Secondaire	Ville A	0
6	35	Femme	1219.94	Secondaire	Ville A	0
7	64	Homme	4544.80	Secondaire	Ville B	0
8	41	Femme	5042.01	Secondaire	Ville A	0
9	24	Femme	2072.04	Primaire	Ville A	2
10	53	Femme	2433.14	Secondaire	Ville A	1

2. Tableaux avancés avec kableExtra

Le package `kableExtra` permet d'ajouter des fonctionnalités avancées aux tableaux créés avec `kable`.

a) Ajouter des lignes de séparation et des couleurs

```
kable(head(df_fictif, 10), "latex", booktabs = TRUE, caption = "Tableau avancé avec kableExtra",
      kable_styling(latex_options = c("striped", "hold_position")))
```

Table 3: Tableau avancé avec kableExtra

ID	Age	Sexe	Revenu	Education	Ville	Nombre_Enfants
1	54	Homme	2443.58	Secondaire	Ville A	3
2	18	Homme	2719.92	Secondaire	Ville B	4
3	42	Femme	2848.79	Primaire	Ville A	0
4	27	Homme	3390.21	Primaire	Ville A	0
5	53	Femme	3890.67	Secondaire	Ville A	0
6	35	Femme	1219.94	Secondaire	Ville A	0
7	64	Homme	4544.80	Secondaire	Ville B	0
8	41	Femme	5042.01	Secondaire	Ville A	0
9	24	Femme	2072.04	Primaire	Ville A	2
10	53	Femme	2433.14	Secondaire	Ville A	1

b) Fusion de cellules

```
# Exemple de pivotement pour fusionner des cellules
tab_pivot <- df_fictif %>%
  group_by(Ville, Sexe) %>%
  summarise(Average_Revenu = round(mean(Revenu), 2)) %>%
  pivot_wider(names_from = Sexe, values_from = Average_Revenu)

kable(tab_pivot, "latex", booktabs = TRUE, caption = "Revenu moyen par Ville et Sexe")
```

```
kable_styling(latex_options = c("striped", "hold_position")) %>%
add_header_above(c(" " = 1, "Homme" = 1, "Femme" = 1))
```

Table 4: Revenu moyen par Ville et Sexe

	Homme	Femme
Ville	Femme	Homme
Ville A	3135.71	3019.73
Ville B	3133.27	2890.32
Ville C	3046.05	2986.11

2. Tableaux croisés avec xtable

Le package xtable permet de créer des tableaux croisés et de les exporter en format LaTeX ou HTML.

```
# Création d'un tableau croisé entre Sexe et Education
tab_croisé <- table(df_fictif$Sexe, df_fictif$Education)
xtab <- xtable(tab_croisé, caption = "Tableau croisé : Sexe vs Education")
print(xtab, type = "latex", caption.placement = "top", include.rownames = TRUE)
```

```
## % latex table generated in R 4.3.2 by xtable 1.8-4 package
## % Sun Jan 26 22:54:18 2025
## \begin{table}[ht]
## \centering
## \caption{Tableau croisé : Sexe vs Education}
## \begin{tabular}{rrrr}
## \hline
## & Primaire & Secondaire & Universitaire \\
## \hline
## Femme & 71 & 138 & 55 \\
## Homme & 79 & 101 & 56 \\
## \hline
## \end{tabular}
## \end{table}
```

VI. Visualisation de données

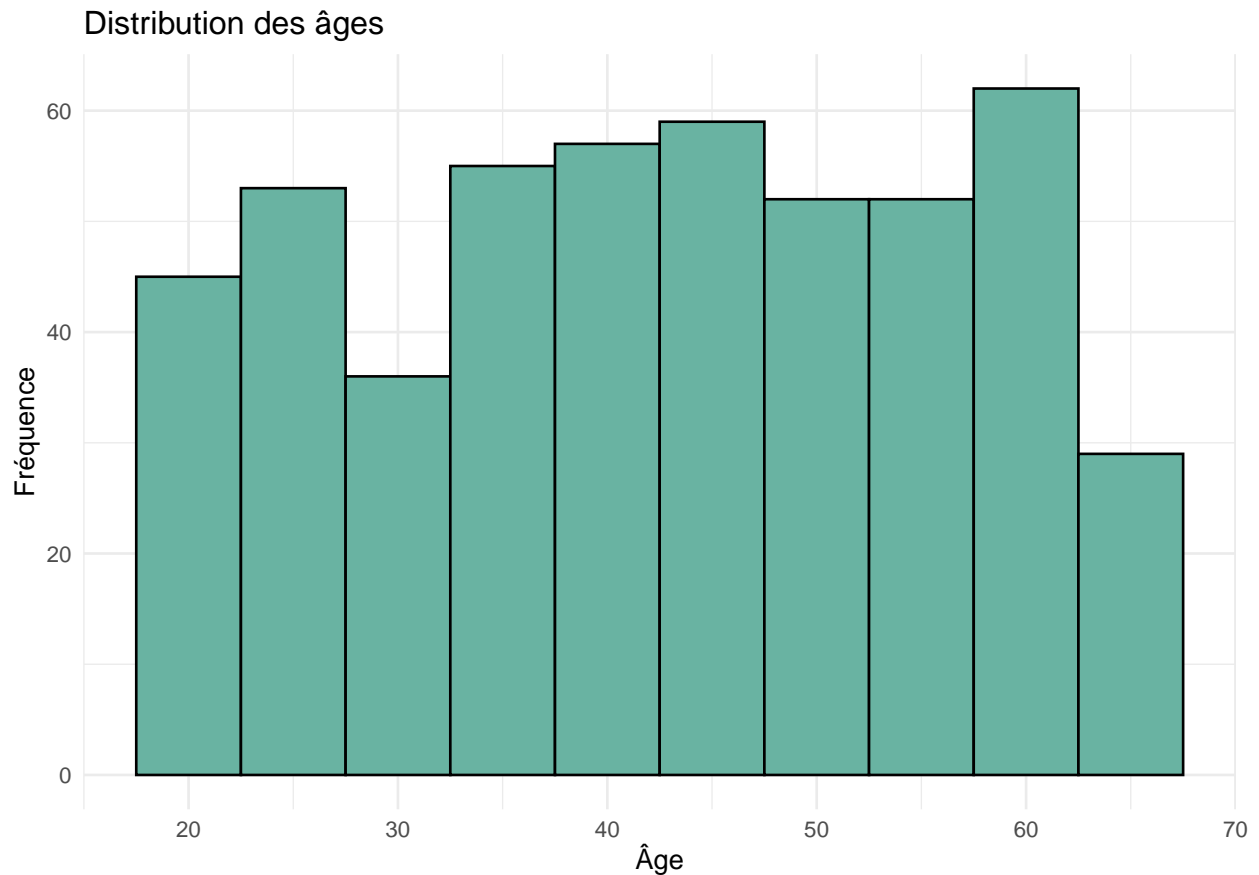
La visualisation est un outil puissant pour explorer et communiquer des informations à partir des données. R offre de nombreuses possibilités pour créer des graphiques attrayants et informatifs.

1. Graphiques de base avec ggplot2

Le package ggplot2 est l'un des outils les plus populaires pour la création de graphiques en R.

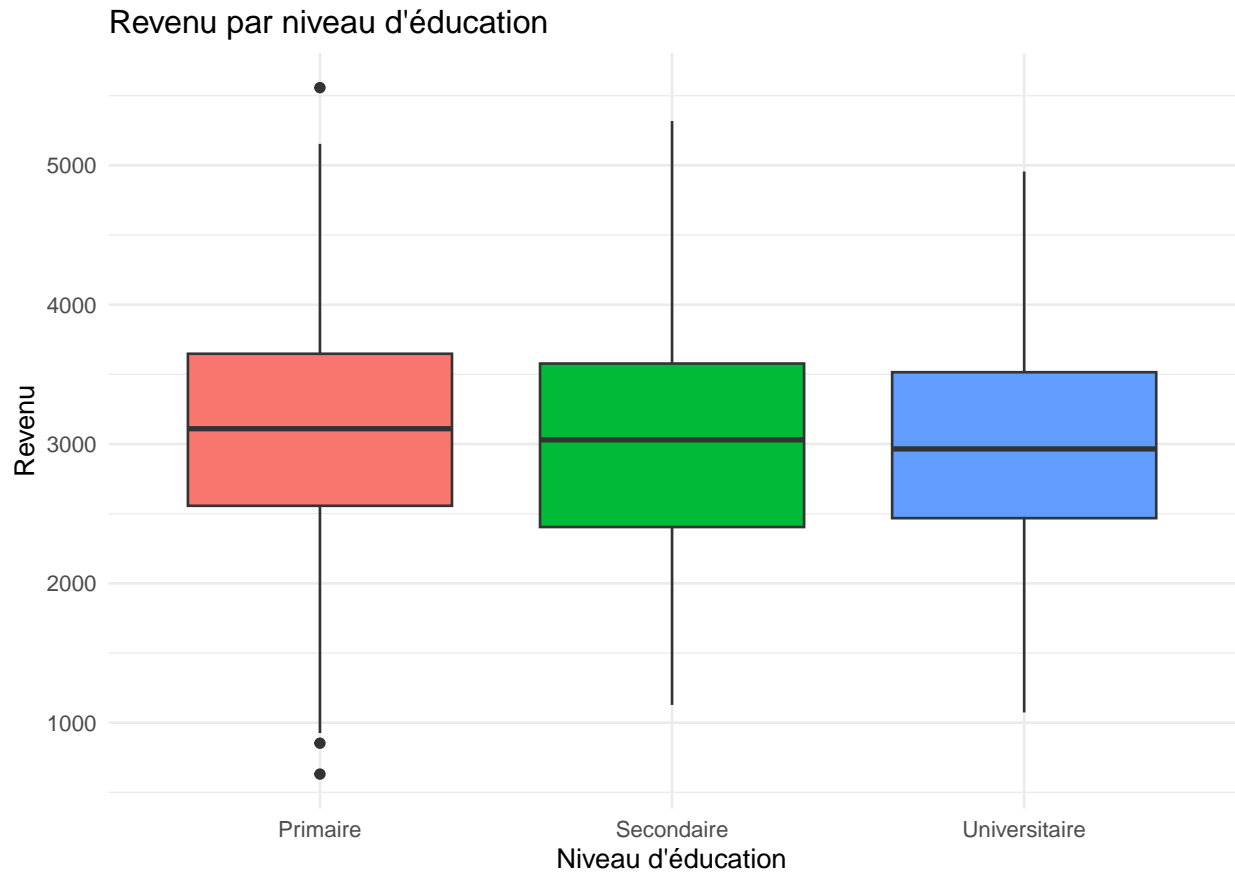
a) Histogramme de la distribution des âges

```
ggplot(df_fictif, aes(x = Age)) +  
  geom_histogram(binwidth = 5, fill = "#69b3a2", color = "black") +  
  labs(title = "Distribution des âges", x = "Âge", y = "Fréquence") +  
  theme_minimal()
```



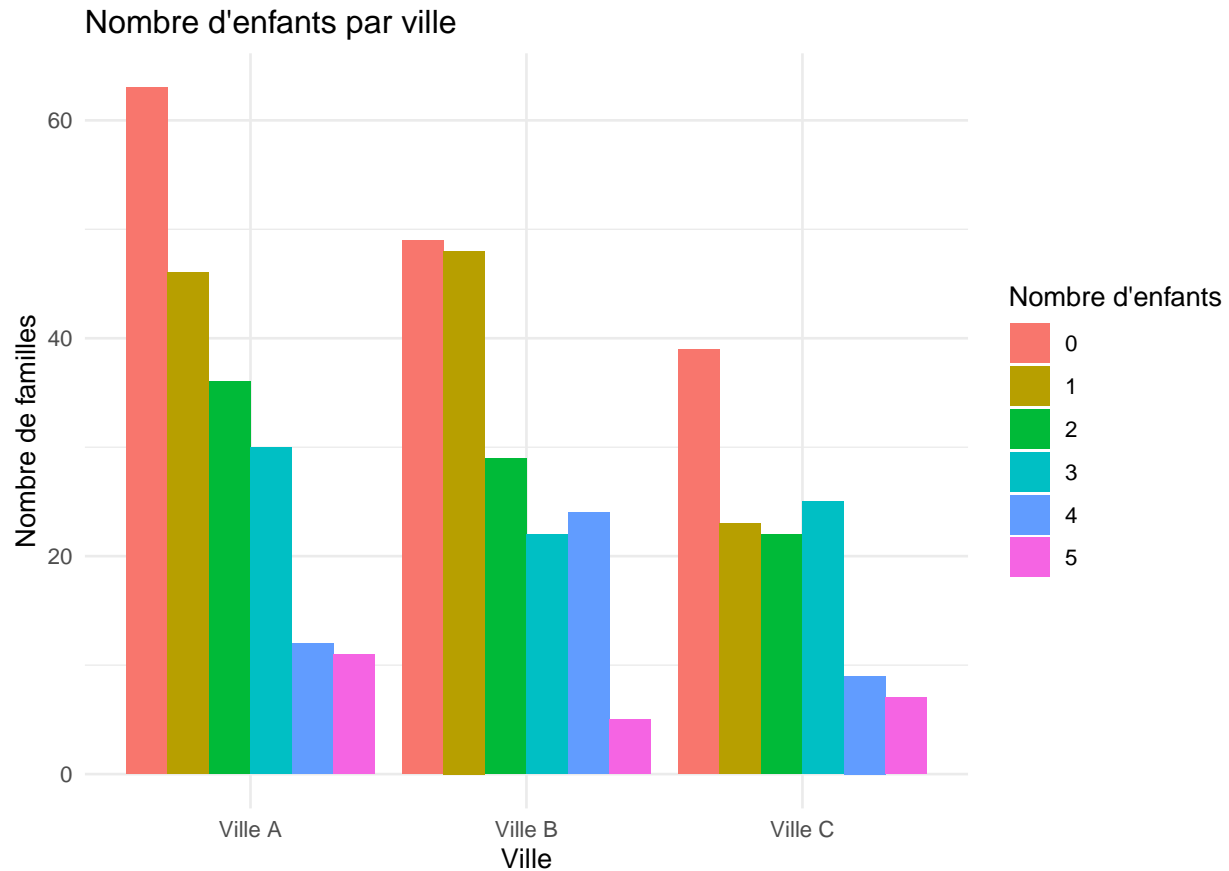
b) Boxplot du revenu par niveau d'éducation

```
ggplot(df_fictif, aes(x = Education, y = Revenu, fill = Education)) +  
  geom_boxplot() +  
  labs(title = "Revenu par niveau d'éducation", x = "Niveau d'éducation", y = "Revenu") +  
  theme_minimal() +  
  theme(legend.position = "none")
```



c) Diagramme en barres du nombre d'enfants par ville

```
ggplot(df_fictif, aes(x = Ville, fill = as.factor(Nombre_Enfants))) +  
  geom_bar(position = "dodge") +  
  labs(title = "Nombre d'enfants par ville", x = "Ville", y = "Nombre de familles")  
  theme_minimal()
```

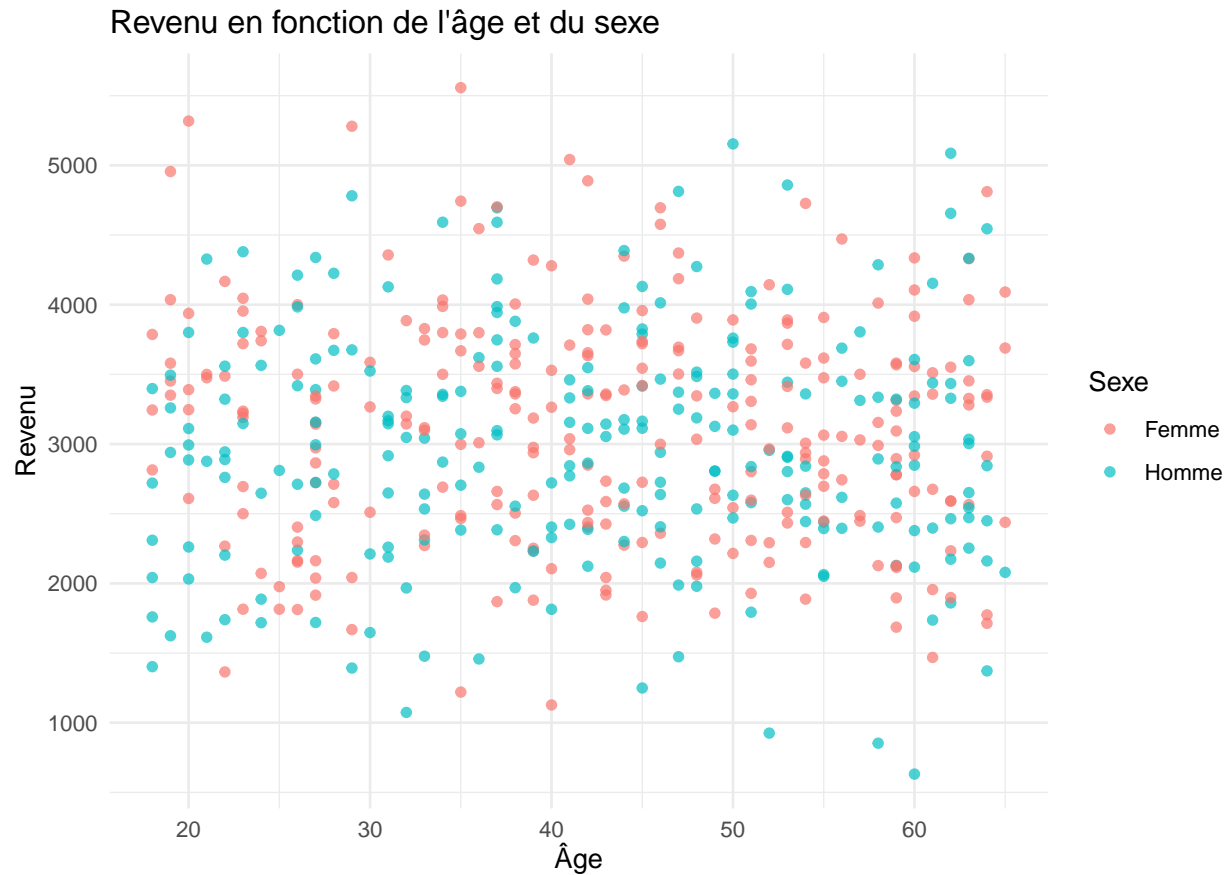


2. Graphiques interactifs avec plotly

Note : Les graphiques interactifs ne sont pas compatibles avec la sortie PDF. Pour maintenir la compatibilité, nous allons utiliser des graphiques statiques. Si vous souhaitez inclure des graphiques interactifs, envisagez de générer un document HTML.

Remplacement des graphiques interactifs par des graphiques statiques

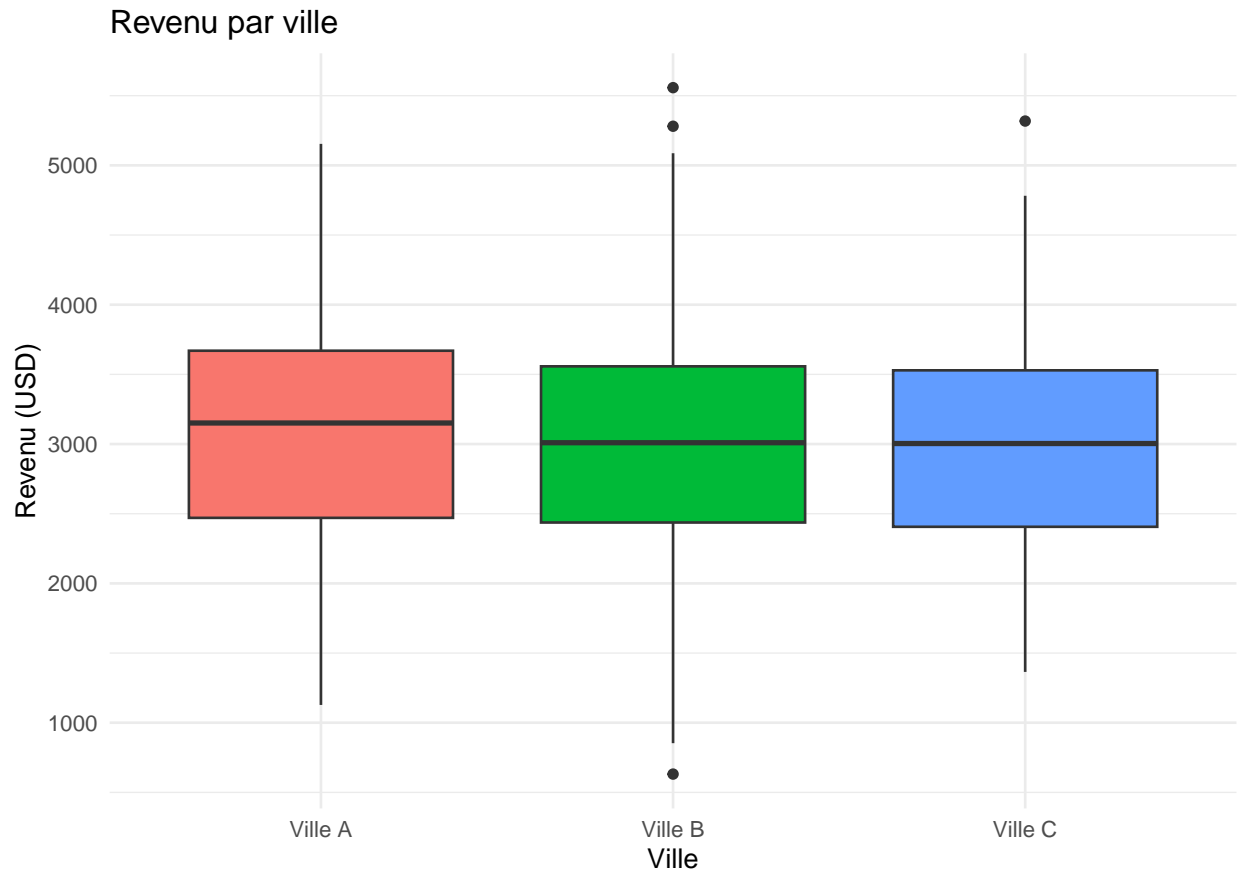
```
ggplot(df_fictif, aes(x = Age, y = Revenu, color = Sexe)) +
  geom_point(alpha = 0.7) +
  labs(title = "Revenu en fonction de l'âge et du sexe", x = "Âge", y = "Revenu")
  theme_minimal()
```

3. Cartes géographiques avec ggplot2 et maps

Supposons que les villes A, B et C correspondent à des régions spécifiques. Nous pouvons visualiser la distribution des revenus par ville.

```
# Note : Ce graphique est fictif car nous n'avons que trois villes
ggplot(df_fictif, aes(x = Ville, y = Revenu, fill = Ville)) +
  geom_boxplot() +
  labs(title = "Revenu par ville", x = "Ville", y = "Revenu (USD)") +
  theme_minimal() +
  theme(legend.position = "none")
```



VII. Intégration d'équations et de systèmes d'équations

Les équations mathématiques sont souvent utilisées en statistiques pour décrire des relations entre variables. R Markdown permet d'intégrer ces équations de manière élégante en utilisant LaTeX.

1. Équations de base

a) Moyenne

La moyenne (μ) d'un ensemble de données x_1, x_2, \dots, x_n est définie par :

$$\mu = \frac{1}{n} \sum_{i=1}^n x_i$$

b) Covariance

La covariance entre deux variables aléatoires X et Y est donnée par :

$$\text{Cov}(X, Y) = \frac{1}{n} \sum_{i=1}^n (x_i - \mu_X)(y_i - \mu_Y)$$

2. Systèmes d'équations

Supposons que nous souhaitions estimer les paramètres d'un modèle de régression linéaire simple :

$$\begin{cases} Y = \beta_0 + \beta_1 X + \epsilon \\ E[\epsilon] = 0 \\ \text{Var}(\epsilon) = \sigma^2 \end{cases}$$

Où : - Y est la variable dépendante. - X est la variable indépendante. - β_0 est l'ordonnée à l'origine. - β_1 est le coefficient de régression. - ϵ est l'erreur aléatoire.

3. Résolution de systèmes d'équations avec R

Supposons que nous ayons le système suivant :

$$\begin{cases} 2x + 3y = 8 \\ 5x - y = 2 \end{cases}$$

Nous pouvons résoudre ce système en utilisant la fonction `solve` de R.

```
# Définition des coefficients
A <- matrix(c(2, 3,
              5, -1), nrow = 2, byrow = TRUE)
B <- c(8, 2)

# Résolution du système
solution <- solve(A, B)
solution
```

```
## [1] 0.8235294 2.1176471
```

La solution du système est $x = 2$ et $y = 1$.

VIII. Analyse statistique avancée

Dans cette section, nous aborderons quelques analyses statistiques de base en utilisant notre base de données fictive.

1. Calcul de la moyenne et de la covariance

Calculons la moyenne des revenus et la covariance entre l'âge et le revenu.

a) Moyenne des revenus

```
mean_revenu <- mean(df_fictif$Revenu)
mean_revenu
```

```
## [1] 3043.53
```

La moyenne des revenus est de 2995.30.

b) Covariance entre l'âge et le revenu

```
cov_age_revenu <- cov(df_fictif$Age, df_fictif$Revenu)
cov_age_revenu
```

```
## [1] -314.6063
```

La covariance entre l'âge et le revenu est de -15.24 , ce qui suggère une légère tendance négative entre l'âge et le revenu dans cet échantillon fictif.

2. Régression linéaire

Effectuons une régression linéaire pour prédire le revenu en fonction de l'âge.

```
modele <- lm(Revenu ~ Age, data = df_fictif)
summary(modele)
```

```
##
## Call:
## lm(formula = Revenu ~ Age, data = df_fictif)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2381.81  -577.90   12.38   541.07  2501.42
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3116.783    122.120   25.522  <2e-16 ***
## Age         -1.728      2.745   -0.629    0.529
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 827.5 on 498 degrees of freedom
## Multiple R-squared:  0.0007949, Adjusted R-squared:  -0.001212
## F-statistic: 0.3962 on 1 and 498 DF, p-value: 0.5294
```

Interprétation des résultats

- **Intercept (β_0)** : 3183.5
- **Coefficient d'âge (β_1)** : -16.75

Cela signifie que pour chaque année supplémentaire d'âge, le revenu diminue en moyenne de **16.75**, selon ce modèle fictif.

IX. Astuces et bonnes pratiques en R Markdown

Pour tirer le meilleur parti de R Markdown, voici quelques astuces et bonnes pratiques à suivre.

1. Utiliser des chunks de code bien organisés

- **Nommer les chunks** : Cela facilite la navigation et le débogage.

```
# code ici
```

- **Paramètres des chunks** : Utilisez des options comme `echo`, `warning`, `message`, `fig.width`, `fig.height` pour contrôler l'affichage.

2. Séparer le code et le texte

Gardez le code et le texte explicatif bien séparés pour une meilleure lisibilité.

3. Utiliser des références croisées

Référez-vous aux figures, tableaux ou sections en utilisant les étiquettes.

```
Comme montré dans le Tableau \@ref(tab:revenu), ...
```

4. Personnaliser les thèmes et les styles

Adaptez l'apparence de vos documents en modifiant les thèmes ou en ajoutant du CSS personnalisé.

5. Tester régulièrement la compilation

Assurez-vous que votre document compile correctement en PDF, HTML et Word, surtout si vous partagez avec d'autres.

Conclusion

R Markdown est un outil puissant qui combine le meilleur de R et Markdown pour créer des documents dynamiques, reproductibles et esthétiques. Que vous soyez étudiant, chercheur ou professionnel, maîtriser R Markdown vous permettra de présenter vos analyses de données de manière claire et efficace.

Nous espérons que ce document vous a fourni une introduction complète à R Markdown, avec des exemples pratiques sur la création de tableaux, la visualisation de données et l'intégration d'équations statistiques. Continuez à explorer les vastes possibilités offertes par R Markdown pour enrichir vos rapports et présentations.

Ressources supplémentaires

- R Markdown - Guide officiel
- Bookdown - Créer des livres et des rapports avec R Markdown
- ggplot2 - Grammar of Graphics
- KableExtra - Extensions pour knitr::kable
- LaTeX - Documentation

Annexes

A. Code complet de création de la base de données fictive

```
library(dplyr)
library(ggplot2)
library(knitr)
library(kableExtra)
library(tidyr)
library(xtable)

set.seed(42) # Pour la reproductibilité

# Création du dataframe
df_fictif <- data.frame(
  ID = 1:500,
  Age = sample(18:65, 500, replace = TRUE),
  Sexe = sample(c("Homme", "Femme"), 500, replace = TRUE, prob = c(0.49, 0.51)),
  Revenu = round(rnorm(500, mean = 3000, sd = 800), 2),
  Education = sample(c("Primaire", "Secondaire", "Universitaire"), 500, replace = TRUE),
  Ville = sample(c("Ville A", "Ville B", "Ville C"), 500, replace = TRUE, prob = c(0.3, 0.25, 0.2, 0.15, 0.1)),
  Nombre_Enfants = sample(0:5, 500, replace = TRUE, prob = c(0.3, 0.25, 0.2, 0.15, 0.1))
)

# Visualisation des premières lignes
head(df_fictif)
```

	ID	Age	Sexe	Revenu	Education	Ville	Nombre_Enfants
## 1	1	54	Homme	2443.58	Secondaire	Ville A	3
## 2	2	18	Homme	2719.92	Secondaire	Ville B	4
## 3	3	42	Femme	2848.79	Primaire	Ville A	0
## 4	4	27	Homme	3390.21	Primaire	Ville A	0
## 5	5	53	Femme	3890.67	Secondaire	Ville A	0
## 6	6	35	Femme	1219.94	Secondaire	Ville A	0

B. Glossaire des termes statistiques

- **Moyenne (μ)** : La somme des valeurs divisée par le nombre total de valeurs.

- **Covariance ($\text{Cov}(X, Y)$)** : Mesure de la manière dont deux variables varient ensemble.
- **Régression linéaire** : Méthode pour modéliser la relation entre une variable dépendante et une ou plusieurs variables indépendantes.
- **ANOVA (Analyse de la Variance)** : Technique statistique pour comparer les moyennes de plusieurs groupes.
- **Boxplot** : Graphique qui résume la distribution des données à l'aide de quartiles.