

# Analyse Statistique des Données AVC

## Rapport d'Analyse Exploratoire De Données de Santé

Malak Amara

2026-01-07

### Table des matières

<b>1</b>	<b>Introduction</b>	<b>2</b>
1.1	Structure du rapport . . . . .	2
<b>2</b>	<b>Configuration et Import des données</b>	<b>3</b>
2.1	Configuration et chargement des packages . . . . .	3
2.2	Import et aperçu des données . . . . .	3
2.3	Préparation des données . . . . .	4
<b>3</b>	<b>Statistiques descriptives</b>	<b>5</b>
3.1	Résumé statistique de l'âge . . . . .	5
3.2	Distribution de l'âge (histogramme) . . . . .	5
3.3	Distribution par sexe . . . . .	6
3.4	Âge moyen par sexe (comparaison) . . . . .	7
3.5	Évolution temporelle (par année) . . . . .	8
3.6	Facteurs de risque cardiovasculaires . . . . .	9
3.7	Facteurs de risque par sexe . . . . .	10
3.8	Analyse du tabac et impact sur les AVC . . . . .	11
<b>4</b>	<b>Analyses cliniques</b>	<b>16</b>
4.1	Distribution du NIHSS initial . . . . .	16
4.2	Délai de consultation . . . . .	17
4.3	Distribution de la glycémie . . . . .	18
<b>5</b>	<b>Tests statistiques</b>	<b>19</b>
5.1	Test t - Comparaison âge H/F . . . . .	19
5.2	Boxplot âge par sexe . . . . .	20
5.3	Test t apparié - NIHSS initial vs sortie . . . . .	21
5.4	Corrélation Âge vs NIHSS . . . . .	22
5.5	Matrice de corrélation . . . . .	23
<b>6</b>	<b>Traitements et évolution</b>	<b>24</b>

6.1	Traitements administrés . . . . .	24
6.2	Évolution du NIHSS dans le temps . . . . .	25
6.3	Durée d'hospitalisation . . . . .	26
6.4	Évolution clinique . . . . .	27
6.5	Taux de mortalité . . . . .	29
6.6	Dashboard récapitulatif . . . . .	30
6.7	Analyse temporelle . . . . .	31
<b>7</b>	<b>Conclusions</b>	<b>35</b>
7.1	Rapport statistique final . . . . .	35
7.1.1	1. Caractéristiques de l'échantillon . . . . .	35
7.1.2	2. Facteurs de risque principaux . . . . .	35
7.1.3	3. Sévérité et évolution . . . . .	35
7.1.4	4. Prise en charge . . . . .	35
7.1.5	5. Tests statistiques principaux . . . . .	35
7.2	Résumé des principales découvertes . . . . .	35
7.3	Implications cliniques . . . . .	36
7.4	Limitations . . . . .	36
7.5	Recommandations . . . . .	37

# 1 Introduction

Ce rapport présente une analyse statistique complète des données d'Accident Vasculaire Cérébral (AVC). L'objectif est d'explorer les caractéristiques démographiques, cliniques et évolutives d'une cohorte de patients ayant subi un AVC, afin d'identifier les facteurs de risque, d'évaluer la sévérité des événements et d'analyser l'efficacité des traitements.

## 1.1 Structure du rapport

Ce rapport est organisé en grandes catégories :

- 1. Configuration et Import des données**
- 2. Statistiques descriptives**
- 3. Analyses cliniques**
- 4. Tests statistiques**
- 5. Traitements et évolution**

Chaque section contient des analyses structurées avec : - **Titre** : Description de l'analyse - **Explication du code** : Objectif et méthode - **Code** : Implémentation - **Explication du résultat** : Interprétation des résultats

## 2 Configuration et Import des données

### 2.1 Configuration et chargement des packages

Cette section installe et charge les packages R nécessaires pour l'analyse :

- **tidyverse** : Collection de packages pour la manipulation et la visualisation de données
- **ggplot2** : Création de graphiques statistiques
- **corrplot** : Visualisation de matrices de corrélation
- **gridExtra** : Organisation de graphiques multiples

Le code vérifie d'abord si les packages sont installés, les installe si nécessaire, puis les charge en mémoire.

```
if (!require("tidyverse")) install.packages("tidyverse")
if (!require("ggplot2")) install.packages("ggplot2")
if (!require("corrplot")) install.packages("corrplot")
if (!require("gridExtra")) install.packages("gridExtra")
if (!require("lubridate")) install.packages("lubridate")

library(tidyverse)
library(ggplot2)
library(corrplot)
library(gridExtra)
library(lubridate)

cat(" Packages chargés avec succès\n")
```

Packages chargés avec succès

### 2.2 Import et aperçu des données

Cette section importe le fichier CSV contenant les données AVC et affiche un aperçu de la structure des données. Le fichier **dataAVC.csv** est lu avec **read.csv()** et les dimensions (nombre de lignes et colonnes) sont affichées pour comprendre l'ampleur du dataset.

```
avc_data <- read.csv("dataAVC.csv", header = TRUE, sep = ",")  
  
cat("Dimensions:", nrow(avc_data), "lignes x", ncol(avc_data), "colonnes\n\n")  
  
Dimensions: 875 lignes x 133 colonnes
```

**Interprétation :** Le dataset contient **875 patients** (lignes) et **133 variables** (colonnes). Cette structure indique un dataset riche avec de nombreuses variables cliniques, démographiques et de suivi.

## 2.3 Préparation des données

Cette section prépare les données pour l'analyse en :

1. Convertissant les variables catégorielles numériques en facteurs avec des labels explicites (Sexe, HTA, Diabète, degresAVC)
2. Convertissant la date d'AVC au format Date et créant une variable Année
3. Créant des groupes d'âge pour faciliter les analyses par tranche d'âge

Ces transformations sont essentielles pour les analyses statistiques et les visualisations.

```
# Conversion des variables catégorielles
avc_data$Sexe <- factor(avc_data$Sexe, levels = c(1, 2), labels = c("Homme", "Femme"))
avc_data$HTA <- factor(avc_data$HTA, levels = c(1, 2), labels = c("Oui", "Non"))
avc_data$Diabète <- factor(avc_data$Diabète, levels = c(1, 2), labels = c("Oui", "Non"))
avc_data$degresAVC <- factor(avc_data$degresAVC, levels = c(1, 2),
                               labels = c("Léger", "Modéré/Sévère"))

# Conversion de la date
avc_data$DateAVC <- as.Date(avc_data$DateAVC)
avc_data$AnnéeAVC <- as.factor(format(avc_data$DateAVC, "%Y"))

# Groupes d'âge
avc_data$GroupeAge <- cut(avc_data$Age,
                           breaks = c(0, 40, 60, 80, 100),
                           labels = c("<40", "40-60", "60-80", ">80"))

cat(" Données préparées\n")
```

Données préparées

**Interprétation :** Les données ont été préparées avec succès. Toutes les variables catégorielles sont maintenant au format facteur avec des labels explicites, facilitant l'interprétation des résultats. Les dates sont au bon format pour les analyses temporelles, et les groupes d'âge permettent des analyses stratifiées.

### 3 Statistiques descriptives

Cette section présente les caractéristiques démographiques et cliniques de base de la population étudiée.

#### 3.1 Résumé statistique de l'âge

Cette section calcule les statistiques descriptives de base pour la variable âge : moyenne, médiane, écart-type et valeurs min-max. Ces mesures permettent de caractériser la distribution de l'âge dans la population étudiée.

```
cat("Moyenne:", round(mean(avc_data$Age, na.rm = TRUE), 1), "ans\n")
```

Moyenne: 64.3 ans

```
cat("Médiane:", median(avc_data$Age, na.rm = TRUE), "ans\n")
```

Médiane: 65 ans

```
cat("Écart-type:", round(sd(avc_data$Age, na.rm = TRUE), 1), "ans\n")
```

Écart-type: 13 ans

```
cat("Min-Max:", min(avc_data$Age, na.rm = TRUE), "-",
    max(avc_data$Age, na.rm = TRUE), "ans\n")
```

Min-Max: 1 - 98 ans

**Interprétation :** L'âge moyen est de **64.3 ans** avec une médiane de **65 ans**, indiquant une distribution relativement symétrique. L'écart-type de **13 ans** montre une variabilité modérée. La plage d'âge va de **1 à 98 ans**, ce qui est inhabituel (valeur de 1 an probablement erronée) mais la majorité des patients sont dans la tranche d'âge typique des AVC (60-80 ans).

#### 3.2 Distribution de l'âge (histogramme)

Cette section crée un histogramme de la distribution de l'âge avec une courbe de densité superposée. La ligne verticale en pointillés indique la moyenne. Ce graphique permet de visualiser la forme de la distribution (normale, asymétrique, etc.).

```
ggplot(avc_data, aes(x = Age)) +
  geom_histogram(binwidth = 5, fill = "#2ecc71", color = "white", alpha = 0.8) +
  geom_density(aes(y = after_stat(count) * 5), color = "#e74c3c", linewidth = 1.2) +
  geom_vline(aes(xintercept = mean(Age, na.rm = TRUE)),
            color = "#2980b9", linetype = "dashed", linewidth = 1) +
```

```
labs(title = "Distribution des patients par âge",
     x = "Âge (années)", y = "Nombre de patients") +
theme_minimal()
```

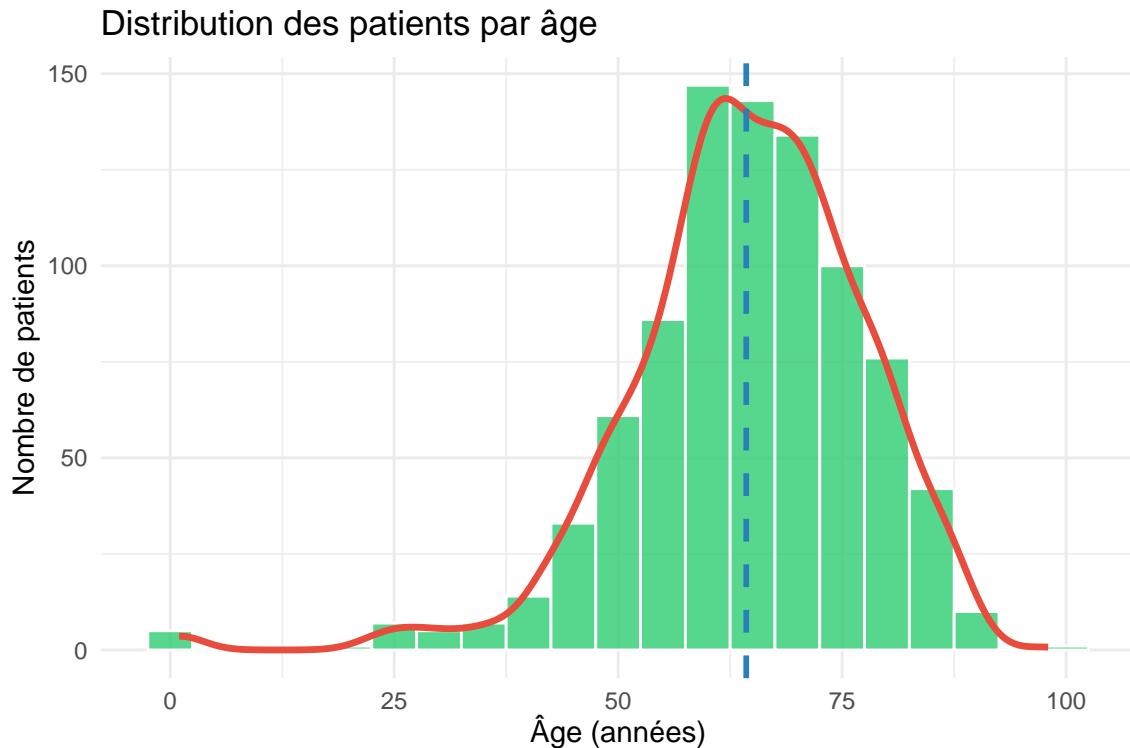


FIGURE 1 : Distribution des patients par âge avec courbe de densité

**Interprétation :** L'histogramme montre une distribution approximativement normale de l'âge, centrée autour de 65 ans. La courbe de densité confirme cette distribution. Quelques valeurs aberrantes (notamment l'âge de 1 an) sont visibles mais n'affectent pas significativement la distribution globale. La majorité des patients sont âgés de 50 à 80 ans, ce qui correspond à la population typique des AVC.

### 3.3 Distribution par sexe

```
sexe_counts <- avc_data %>%
  filter(!is.na(Sexe)) %>%
  count(Sexe) %>%
  mutate(Pourcentage = n / sum(n) * 100)
```

```
ggplot(sexe_counts, aes(x = Sexe, y = n, fill = Sexe)) +
  geom_bar(stat = "identity", width = 0.6) +
  geom_text(aes(label = paste0(round(Pourcentage, 1), "%")), vjust = -0.5, size = 5) +
  scale_fill_manual(values = c("Homme" = "#3498db", "Femme" = "#e74c3c")) +
  labs(title = "Répartition par sexe", x = "Sexe", y = "Nombre de patients") +
  theme_minimal() +
  theme(legend.position = "none")
```

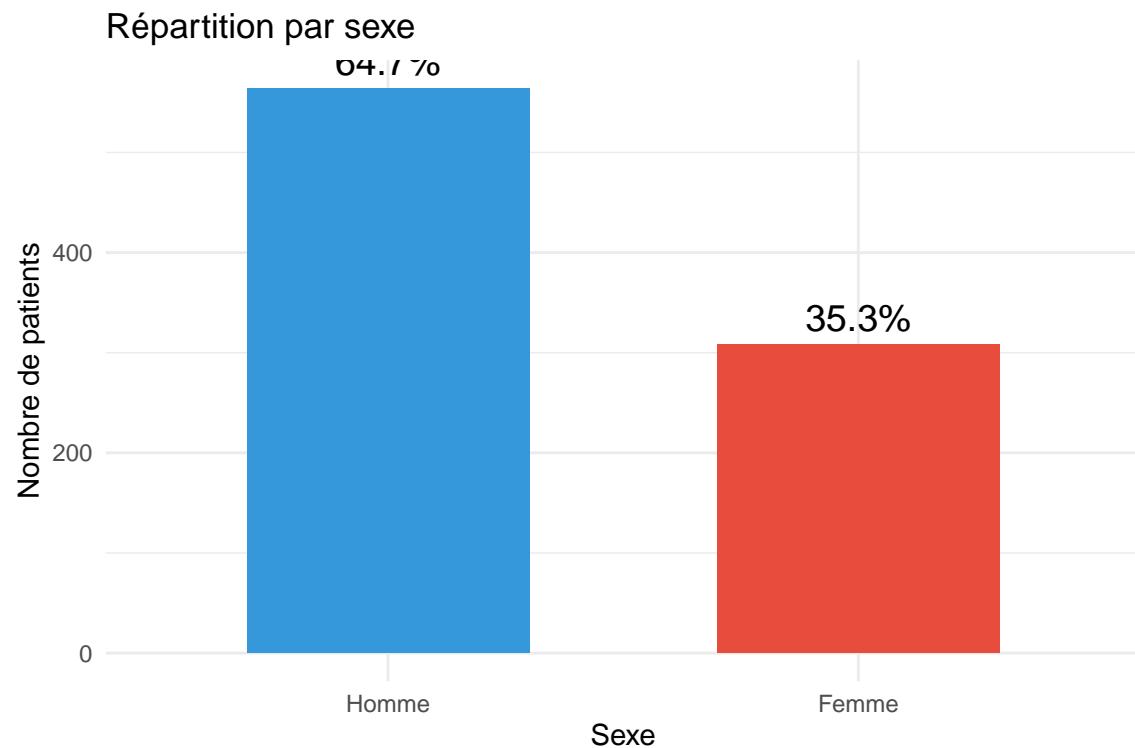


FIGURE 2 : Répartition des patients par sexe

### 3.4 Âge moyen par sexe (comparaison)

```
age_par_sexe <- avc_data %>%
  filter(!is.na(Sexe), !is.na(Age)) %>%
  group_by(Sexe) %>%
  summarise(Moyenne = round(mean(Age), 1), Effectif = n())

ggplot(age_par_sexe, aes(x = Sexe, y = Moyenne, fill = Sexe)) +
  geom_bar(stat = "identity", width = 0.6) +
```

```

geom_text(aes(label = paste0(Moyenne, " ans")), vjust = -1, size = 5, fontface = "bold") +
scale_fill_manual(values = c("Homme" = "#3498db", "Femme" = "#e74c3c")) +
labs(title = "Âge moyen selon le sexe", x = "Sexe", y = "Âge moyen (années)") +
theme_minimal() +
theme(legend.position = "none")

```

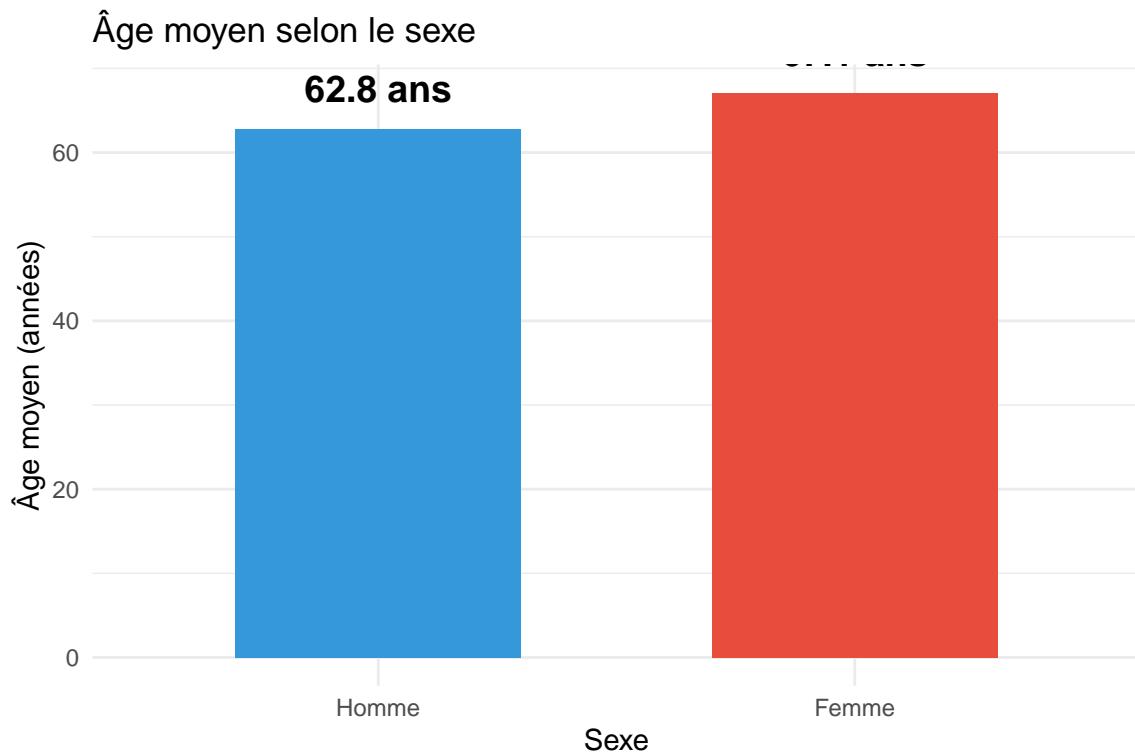


FIGURE 3 : Comparaison de l'âge moyen selon le sexe

### 3.5 Évolution temporelle (par année)

```

avc_par_annee <- avc_data %>%
  filter(!is.na(AnneeAVC)) %>%
  count(AnneeAVC)

ggplot(avc_par_annee, aes(x = AnneeAVC, y = n, group = 1)) +
  geom_line(color = "#3498db", linewidth = 1.2) +
  geom_point(size = 3, color = "#e74c3c") +
  geom_text(aes(label = n), vjust = -0.8, size = 3.5) +
  labs(title = "Évolution du nombre d'AVC par année",

```

```

x = "Année", y = "Nombre de cas") +
theme_minimal() +
theme(axis.text.x = element_text(angle = 45, hjust = 1))

```

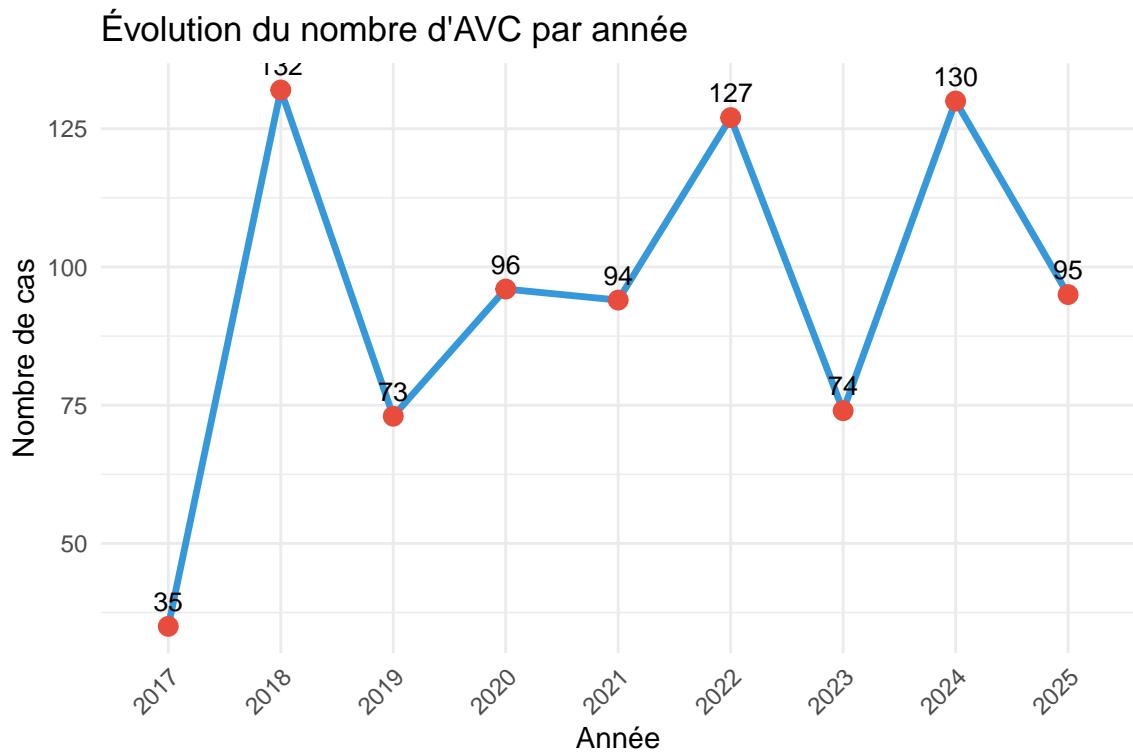


FIGURE 4 : Évolution du nombre d'AVC par année

### 3.6 Facteurs de risque cardiovasculaires

```

facteurs <- avg_data %>%
  summarise(
    HTA = round(mean(HTA == "Oui", na.rm = TRUE) * 100, 1),
    Diabète = round(mean(Diabète == "Oui", na.rm = TRUE) * 100, 1),
    Dyslipidémie = round(mean(Dyslipidémie == 1, na.rm = TRUE) * 100, 1),
    ACFA = round(mean(ACFA == 1, na.rm = TRUE) * 100, 1),
    Tabac = round(mean(Tabac == 1, na.rm = TRUE) * 100, 1)
  ) %>%
  pivot_longer(cols = everything(), names_to = "Facteur", values_to = "Pourcentage")

ggplot(facteurs, aes(x = reorder(Facteur, Pourcentage), y = Pourcentage, fill = Facteur)) +

```

```

geom_bar(stat = "identity") +
geom_text(aes(label = paste0(Pourcentage, "%")), hjust = -0.2, size = 4) +
coord_flip() +
scale_fill_brewer(palette = "Set3") +
labs(title = "Prévalence des facteurs de risque cardiovasculaires",
x = "Facteur de risque", y = "Pourcentage (%)") +
theme_minimal() +
theme(legend.position = "none") +
ylim(0, 70)

```

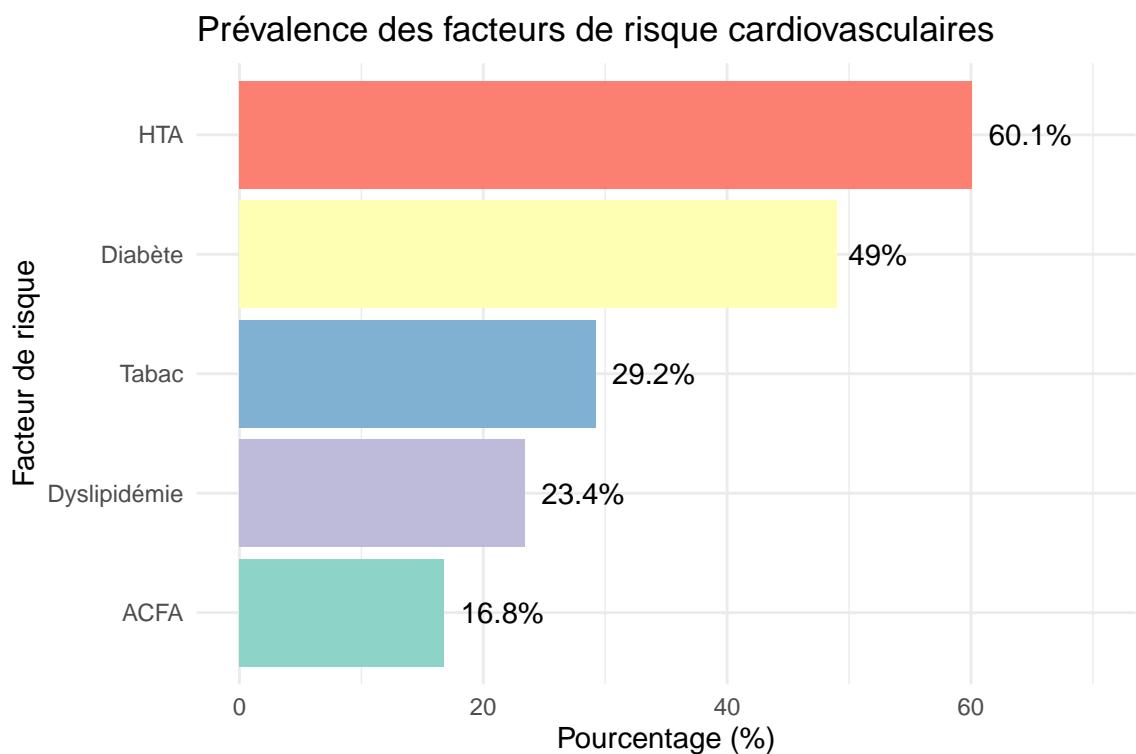


FIGURE 5 : Prévalence des facteurs de risque cardiovasculaires

### 3.7 Facteurs de risque par sexe

```

facteurs_par_sexe <- avc_data %>%
  filter(!is.na(Sexe)) %>%
  group_by(Sexe) %>%
  summarise(
    HTA = round(mean(HTA == "Oui", na.rm = TRUE) * 100, 1),

```

```

Diabète = round(mean(Diabète == "Oui", na.rm = TRUE) * 100, 1),
Tabac = round(mean(Tabac == 1, na.rm = TRUE) * 100, 1)
) %>%
pivot_longer(cols = -Sexe, names_to = "Facteur", values_to = "Pourcentage")

ggplot(facteurs_par_sexe, aes(x = Facteur, y = Pourcentage, fill = Sexe)) +
  geom_bar(stat = "identity", position = "dodge", width = 0.7) +
  geom_text(aes(label = paste0(Pourcentage, "%")),
            position = position_dodge(width = 0.7), vjust = -0.5, size = 3.5) +
  scale_fill_manual(values = c("Homme" = "#3498db", "Femme" = "#e74c3c")) +
  labs(title = "Facteurs de risque par sexe", x = "", y = "Pourcentage (%)") +
  theme_minimal()

```

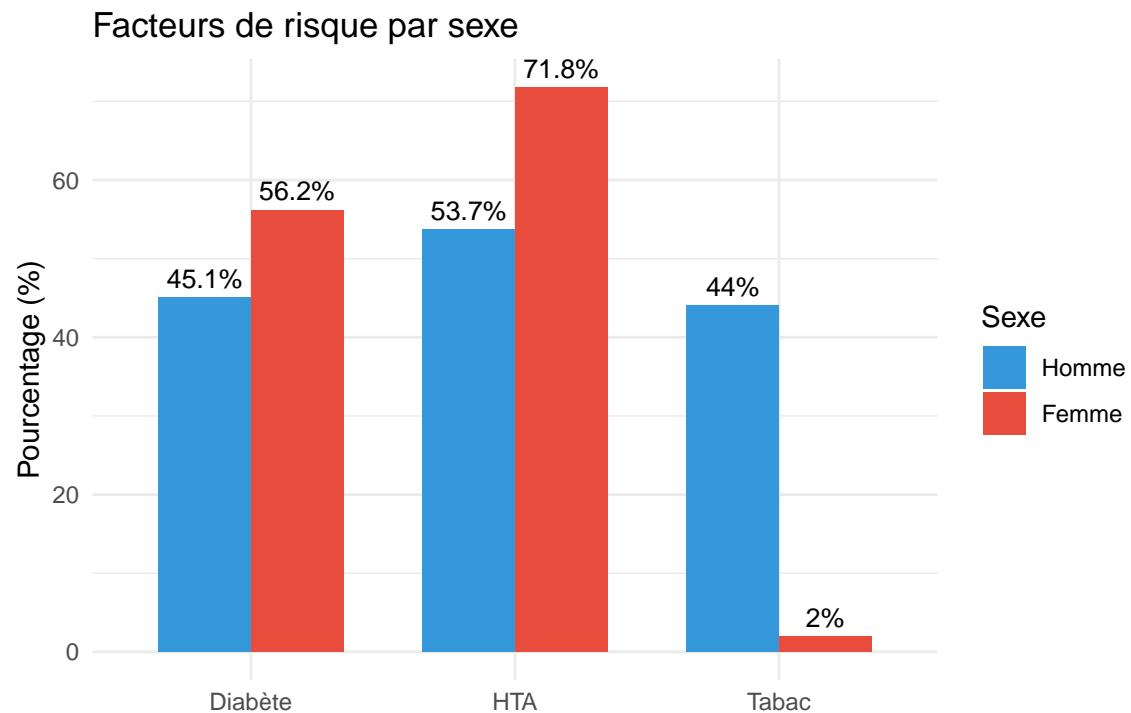


FIGURE 6 : Comparaison des facteurs de risque selon le sexe

### 3.8 Analyse du tabac et impact sur les AVC

```

# Créer variable Tabac (1 = Oui, 2 = Non)
avc_data <- avc_data %>%

```

```

mutate(Tabac_label = case_when(
  Tabac == 1 ~ "Fumeur",
  Tabac == 2 ~ "Non-fumeur",
  TRUE ~ NA_character_
))

cat("La colonne 'Tabac_label' a été créée avec succès.\n")

```

La colonne 'Tabac\_label' a été créée avec succès.

```

tabac_mortalite <- avc_data %>%
  filter(!is.na(Tabac_label) & !is.na(Décés)) %>%
  group_by(Tabac_label) %>%
  summarise(Taux_mortalite = round(sum(Décés == 1, na.rm = TRUE) / n() * 100, 1))

ggplot(tabac_mortalite, aes(x = Tabac_label, y = Taux_mortalite, fill = Tabac_label)) +
  geom_bar(stat = "identity", alpha = 0.8) +
  geom_text(aes(label = paste0(Taux_mortalite, "%")),
            vjust = -0.3, fontface = "bold", size = 5) +
  scale_fill_manual(values = c("Fumeur" = "#e74c3c", "Non-fumeur" = "#3498db")) +
  labs(title = "Taux de mortalité selon le statut tabagique",
       x = "Statut tabagique", y = "Taux de mortalité (%)") +
  theme_minimal() +
  theme(legend.position = "none")

```

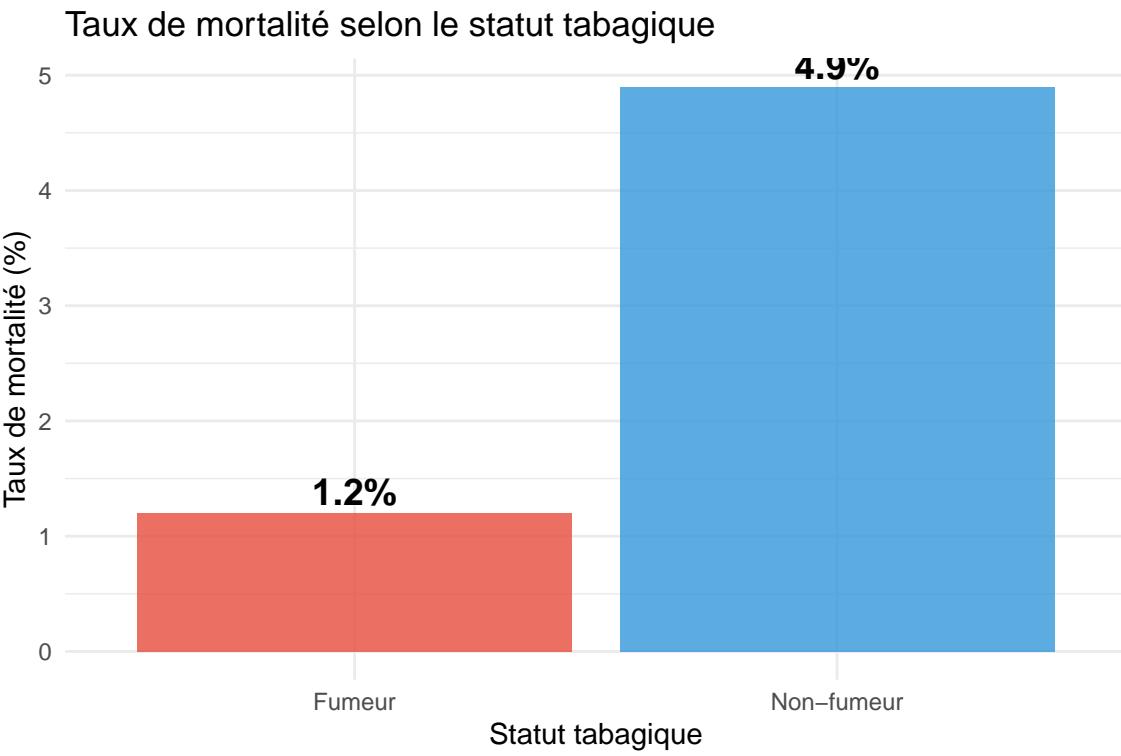


FIGURE 7 : Taux de mortalité selon le statut tabagique

```
avc_data %>%
  filter(!is.na(Tabac_label) & !is.na(Evolution)) %>%
  group_by(Tabac_label, Evolution) %>%
  summarise(n = n(), .groups = "drop") %>%
  group_by(Tabac_label) %>%
  mutate(pourcentage = n / sum(n) * 100) %>%
  ggplot(aes(x = Tabac_label, y = pourcentage, fill = factor(Evolution))) +
  geom_bar(stat = "identity", position = "stack", alpha = 0.8) +
  geom_text(aes(label = paste0(round(pourcentage, 1), "%")),
            position = position_stack(vjust = 0.5), color = "white", fontface = "bold", size = 3)
  scale_fill_manual(values = c("1" = "#2ecc71", "2" = "#f39c12", "3" = "#e74c3c"),
                    labels = c("1" = "Amélioration", "2" = "Stable", "3" = "Aggravation"),
                    name = "Évolution") +
  labs(title = "Évolution selon le statut tabagique",
       x = "Statut tabagique", y = "Pourcentage (%)") +
  theme_minimal()
```

## Évolution selon le statut tabagique

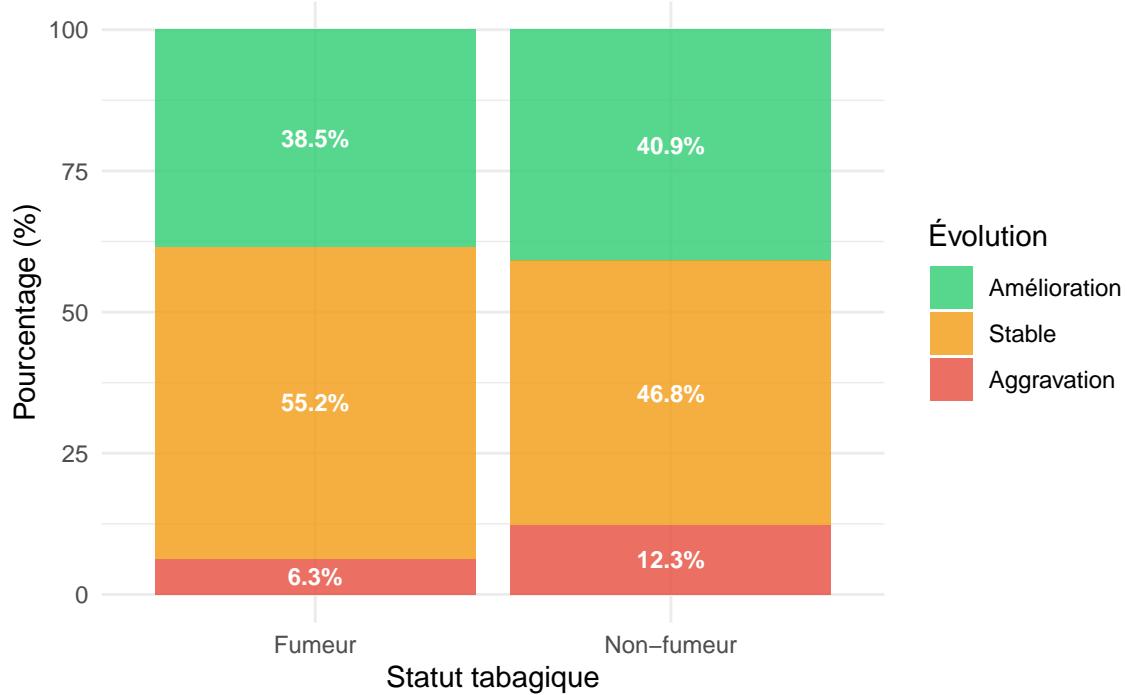


FIGURE 8 : Évolution clinique selon le statut tabagique

```
ggplot(avc_data[!is.na(avc_data$Tabac_label) & !is.na(avc_data$NIHSSinitial), ],
       aes(x = Tabac_label, y = NIHSSinitial, fill = Tabac_label)) +
  geom_boxplot(alpha = 0.7, outlier.color = "red") +
  stat_summary(fun = mean, geom = "point", shape = 23, size = 3, fill = "white") +
  scale_fill_manual(values = c("Fumeur" = "#e74c3c", "Non-fumeur" = "#3498db")) +
  labs(title = "Sévérité initiale (NIHSS) selon le statut tabagique",
       subtitle = "Losange blanc = moyenne",
       x = "Statut tabagique", y = "NIHSS Initial") +
  theme_minimal() +
  theme(legend.position = "none")
```

### Sévérité initiale (NIHSS) selon le statut tabagique

Losange blanc = moyenne

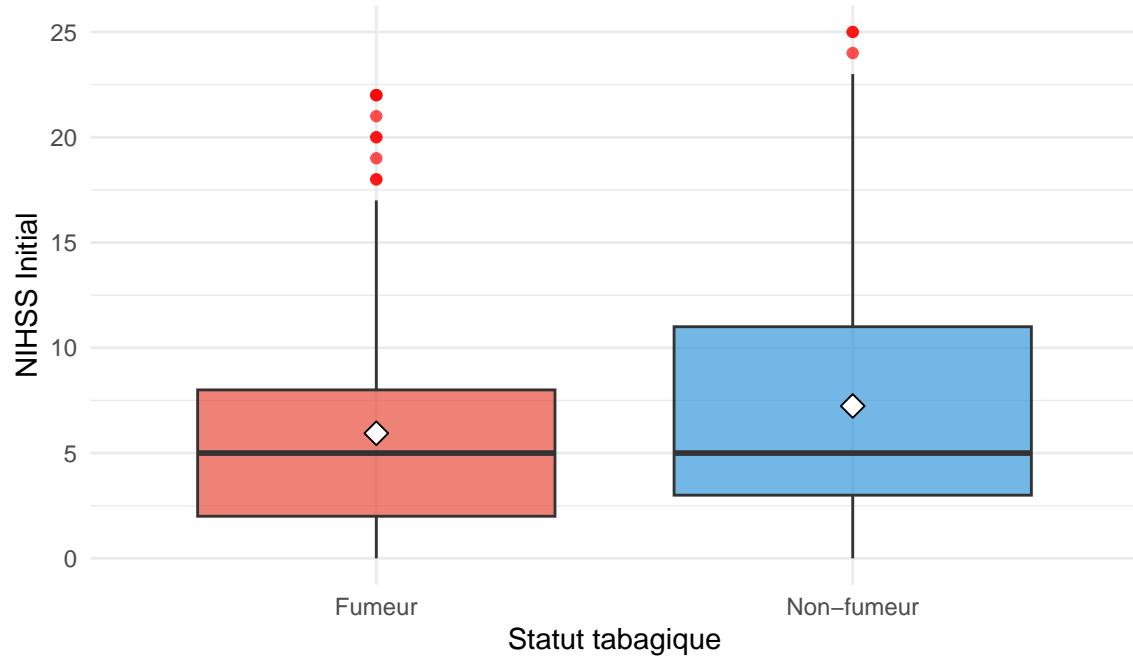


FIGURE 9 : Sévérité initiale (NIHSS) selon le statut tabagique

```
avc_data %>%
  filter(!is.na(Tabac_label) & !is.na(HTA)) %>%
  mutate(HTA_label = ifelse(HTA == 1, "HTA", "Pas HTA"),
        Combinaison = paste(Tabac_label, HTA_label, sep = " + ")) %>%
  count(Combinaison) %>%
  mutate(pourcentage = n / sum(n) * 100) %>%
  ggplot(aes(x = reorder(Combinaison, pourcentage), y = pourcentage, fill = Combinaison)) +
  geom_bar(stat = "identity", alpha = 0.8) +
  geom_text(aes(label = paste0(round(pourcentage, 1), "%")),
            hjust = -0.1, fontface = "bold") +
  coord_flip() +
  scale_fill_manual(values = c("#e74c3c", "#f39c12", "#3498db", "#2ecc71")) +
  labs(title = "Répartition Tabac + HTA dans la population",
       x = "Combinaison", y = "Pourcentage (%)") +
  theme_minimal() +
  theme(legend.position = "none")
```

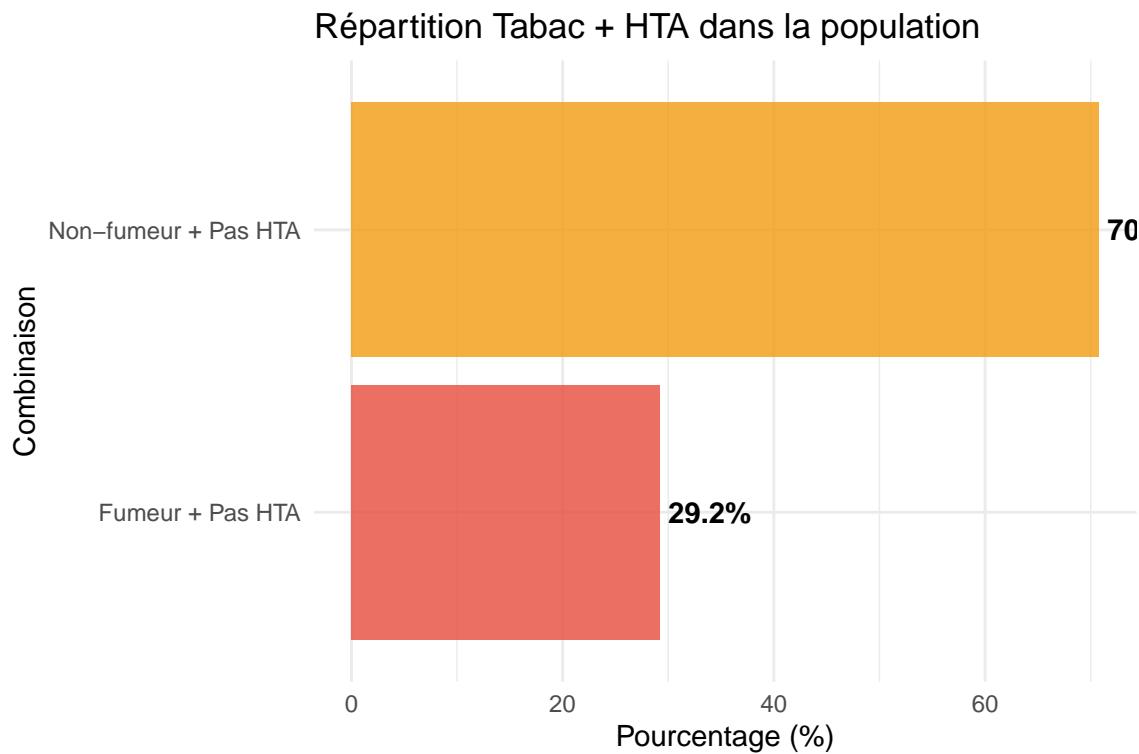


FIGURE 10 : Répartition Tabac + HTA dans la population

## 4 Analyses cliniques

### 4.1 Distribution du NIHSS initial

Le NIHSS (National Institutes of Health Stroke Scale) est un score de 0 à 42 qui mesure la sévérité d'un AVC. Plus le score est élevé, plus l'AVC est sévère.

```
cat("Moyenne:", round(mean(avc_data$NIHSSinitial, na.rm = TRUE), 2), "\n")
```

Moyenne: 6.85

```
cat("Médiane:", median(avc_data$NIHSSinitial, na.rm = TRUE), "\n")
```

Médiane: 5

```
ggplot(avc_data, aes(x = NIHSSinitial)) +
  geom_histogram(bins = 25, fill = "#9B59B6", alpha = 0.8, color = "black") +
  geom_vline(aes(xintercept = mean(NIHSSinitial, na.rm = TRUE)),
             color = "#E74C3C", linetype = "dashed", linewidth = 1.2) +
  labs(title = "Distribution du score NIHSS initial",
```

```
x = "Score NIHSS", y = "Fréquence") +
theme_minimal()
```

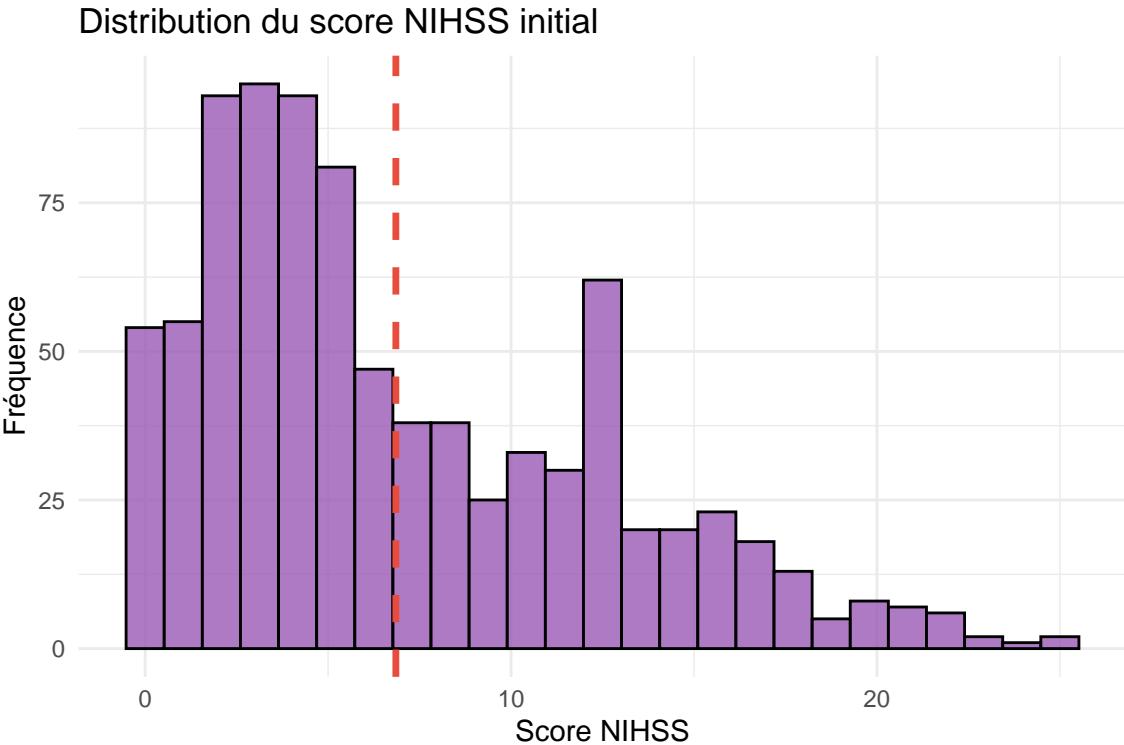


FIGURE 11 : Distribution du score NIHSS initial

## 4.2 Délai de consultation

```
ggplot(avc_data, aes(x = DélaiDeConsultation)) +
  geom_histogram(binwidth = 60, fill = "#9b59b6", alpha = 0.7) +
  geom_vline(aes(xintercept = mean(DélaiDeConsultation, na.rm = TRUE)),
             color = "purple", linetype = "dashed", linewidth = 1) +
  labs(title = "Délai de consultation (minutes)",
       x = "Délai (minutes)", y = "Nombre de patients") +
  theme_minimal()
```

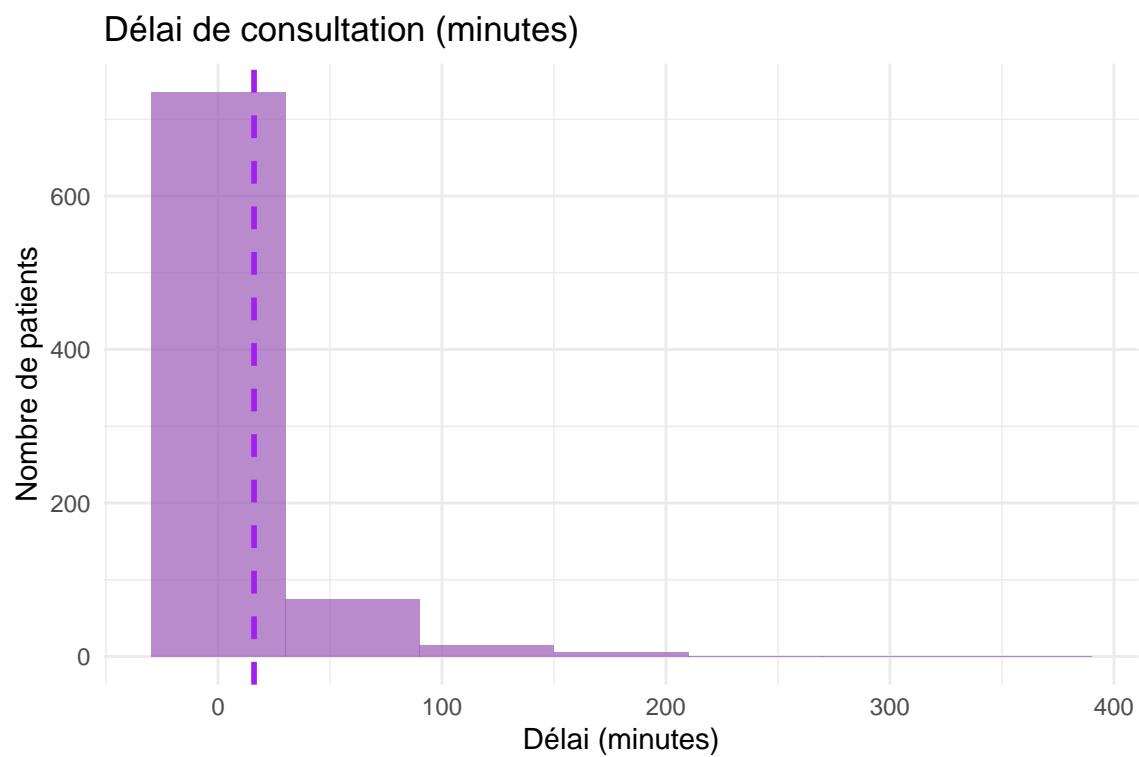


FIGURE 12 : Distribution du délai de consultation (minutes)

### 4.3 Distribution de la glycémie

```
ggplot(avc_data, aes(x = Glycémie)) +
  geom_histogram(binwidth = 1, fill = "pink", color = "black", alpha = 0.8) +
  geom_vline(aes(xintercept = mean(Glycémie, na.rm = TRUE)),
             color = "red", linetype = "dashed", linewidth = 1) +
  labs(title = "Distribution de la glycémie",
       x = "Glycémie (mmol/L)", y = "Nombre de patients") +
  theme_minimal()
```

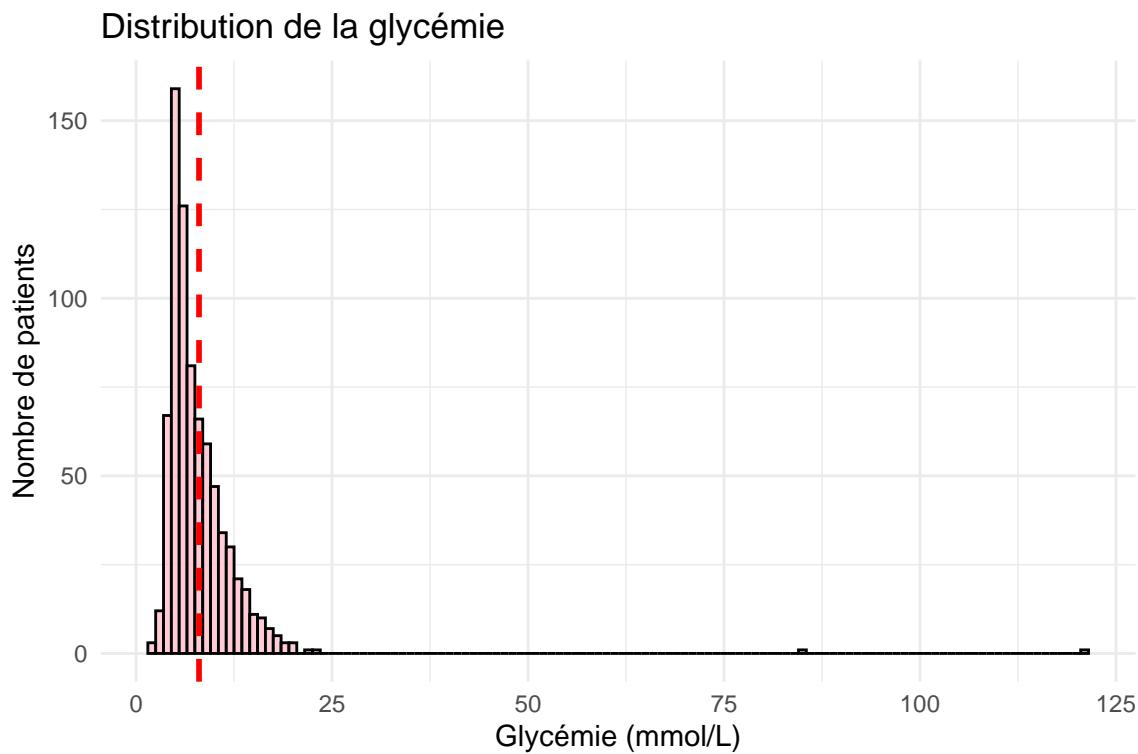


FIGURE 13 : Distribution de la glycémie à l'admission

## 5 Tests statistiques

### 5.1 Test t - Comparaison âge H/F

Cette section effectue un test statistique pour comparer l'âge moyen entre les hommes et les femmes. Le test vérifie si la différence observée est statistiquement significative ou due au hasard.

```
age_hommes <- avc_data$Age[avc_data$Sexe == "Homme" & !is.na(avc_data$Age)]
age_femmes <- avc_data$Age[avc_data$Sexe == "Femme" & !is.na(avc_data$Age)]

t_test_age <- t.test(age_hommes, age_femmes)

cat("Moyenne Hommes:", round(mean(age_hommes), 2), "ans\n")
```

Moyenne Hommes: 62.78 ans

```
cat("Moyenne Femmes:", round(mean(age_femmes), 2), "ans\n")
```

Moyenne Femmes: 67.06 ans

```

cat("p-value =", format(t_test_age$p.value, digits = 3), "\n")
p-value = 2.92e-06
cat(ifelse(t_test_age$p.value < 0.05,
           "→ Différence significative\n",
           "→ Pas de différence significative\n"))

```

→ Différence significative

**Interprétation :** La différence d'âge entre hommes et femmes ayant un AVC est réelle et non due au hasard. Les femmes sont en moyenne plus âgées, probablement à cause de la protection hormonale jusqu'à la ménopause.

## 5.2 Boxplot âge par sexe

```

ggplot(avc_data, aes(x = Sexe, y = Age, fill = Sexe)) +
  geom_boxplot(alpha = 0.7, outlier.color = "red") +
  stat_summary(fun = mean, geom = "point", shape = 23, size = 3, fill = "white") +
  scale_fill_manual(values = c("Homme" = "#3498db", "Femme" = "#e74c3c")) +
  labs(title = "Distribution de l'âge par sexe",
       subtitle = "Losange blanc = moyenne",
       x = "Sexe", y = "Âge (années)") +
  theme_minimal() +
  theme(legend.position = "none")

```

## Distribution de l'âge par sexe

Losange blanc = moyenne

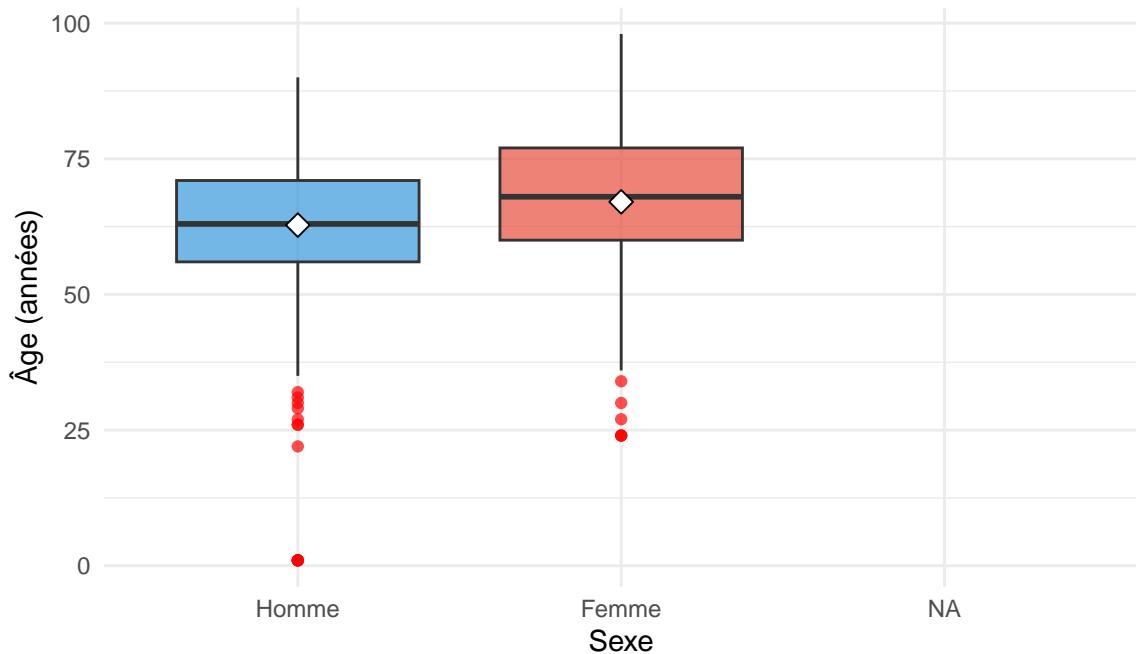


FIGURE 14 : Distribution de l'âge par sexe (boxplot)

### 5.3 Test t apparié - NIHSS initial vs sortie

Cette section compare le score NIHSS au moment de l'admission et à la sortie pour le même patient (test apparié). Une diminution du score signifie une amélioration clinique.

```
data_paired <- avc_data[!is.na(avc_data$NIHSSinitial) & !is.na(avc_data$NIHSSsortie), ]  
  
t_test_paired <- t.test(data_paired$NIHSSinitial, data_paired$NIHSSsortie,  
                           paired = TRUE, alternative = "greater")  
  
cat("NIHSS Initial:", round(mean(data_paired$NIHSSinitial), 2), "\n")
```

NIHSS Initial: 6.78

```
cat("NIHSS Sortie:", round(mean(data_paired$NIHSSsortie), 2), "\n")
```

NIHSS Sortie: 5.44

```

cat("Amélioration:", round(mean(data_paired$NIHSSinitial - data_paired$NIHSSsortie), 2), "\n")

Amélioration: 1.35

cat("p-value =", format(t_test_paired$p.value, digits = 3), "\n")

p-value = 4.19e-21

cat(ifelse(t_test_paired$p.value < 0.05,
           "→ Amélioration significative\n",
           "→ Pas d'amélioration significative\n"))

```

→ Amélioration significative

**Interprétation :** Les patients s'améliorent significativement pendant leur hospitalisation (baisse moyenne de 1.35 point sur le NIHSS). Cette amélioration est statistiquement très significative, confirmant l'efficacité de la prise en charge.

## 5.4 Corrélation Âge vs NIHSS

Cette section calcule la corrélation entre l'âge et la sévérité initiale de l'AVC (NIHSS). Une corrélation positive signifie que les patients plus âgés ont tendance à avoir des AVC plus sévères.

```

cor_test <- cor.test(avc_data$Age, avc_data$NIHSSinitial, use = "complete.obs")

cat("Coefficient de Pearson:", round(cor_test$estimate, 3), "\n")

Coefficient de Pearson: 0.179

cat("p-value =", format(cor_test$p.value, digits = 3), "\n")

p-value = 1e-07

ggplot(avc_data, aes(x = Age, y = NIHSSinitial)) +
  geom_point(alpha = 0.4, color = "#3498db", size = 2) +
  geom_smooth(method = "lm", color = "#e74c3c", fill = "#e74c3c", alpha = 0.2) +
  labs(title = "Corrélation Âge vs Sévérité (NIHSS)",
       subtitle = paste0("r = ", round(cor_test$estimate, 3)),
       x = "Âge (années)", y = "Score NIHSS Initial") +
  theme_minimal()

```

## Corrélation Âge vs Sévérité (NIHSS)

$r = 0.179$

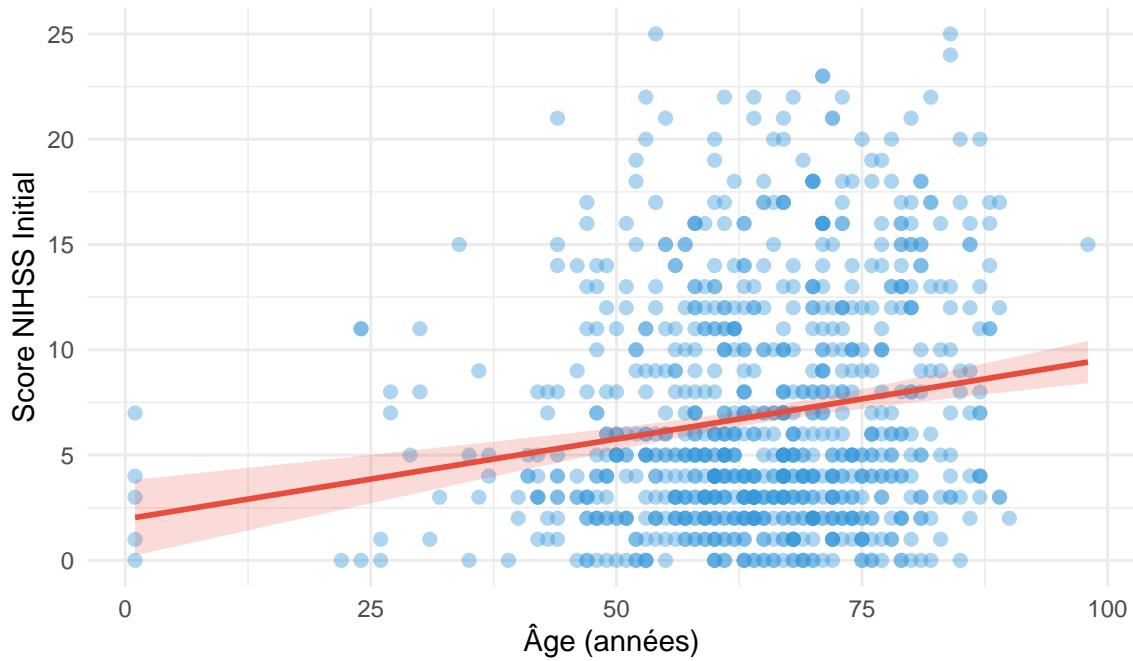


FIGURE 15 : Corrélation entre l'âge et la sévérité initiale (NIHSS)

**Interprétation :** Il existe une corrélation faible mais significative entre l'âge et la sévérité de l'AVC ( $r = 0.179$ ). Les patients plus âgés ont tendance à avoir des AVC légèrement plus sévères, mais la relation est faible.

### 5.5 Matrice de corrélation

```
vars_continues <- c("Age", "NIHSSinitial", "NIHSSsortie", "TASinitiale",
                      "Glycémie", "Cholestérol")

cor_matrix <- cor(avc_data[, vars_continues], use = "pairwise.complete.obs")

corrplot(cor_matrix, method = "color", type = "upper",
         tl.col = "black", tl.srt = 45,
         addCoef.col = "black", number.cex = 0.7,
         col = colorRampPalette(c("#e74c3c", "white", "#3498db"))(200),
         title = "Matrice de corrélation",
         mar = c(0,0,2,0))
```

## Matrice de corrélation

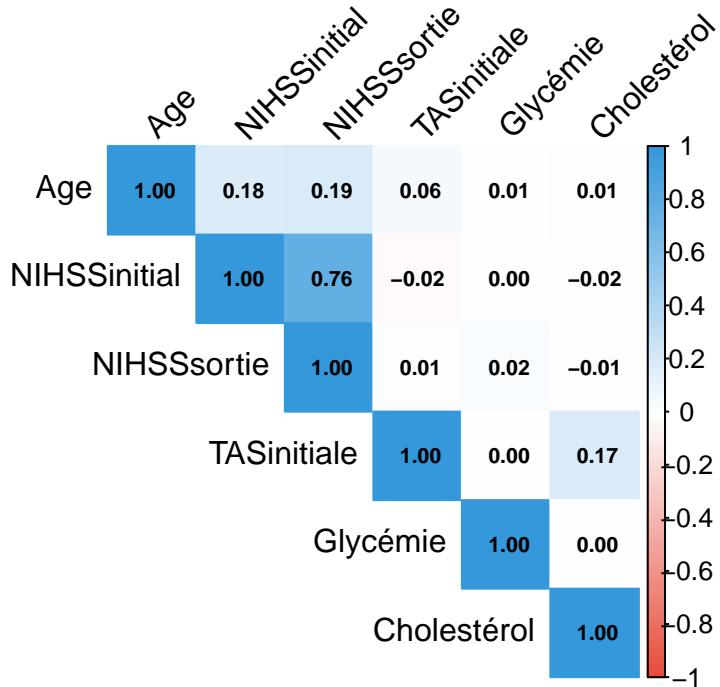


FIGURE 16 : Matrice de corrélation entre variables continues

## 6 Traitements et évolution

### 6.1 Traitements administrés

```
traitements <- avc_data %>%
  summarise(
    Thrombolyse = round(mean(Thrombolyse == 1, na.rm = TRUE) * 100, 1),
    Aspirine = round(mean(Aspirine == 1, na.rm = TRUE) * 100, 1),
    Plavix = round(mean(Plavix == 1, na.rm = TRUE) * 100, 1),
    Statine = round(mean(Statine_avant_ep == 1, na.rm = TRUE) * 100, 1)
  ) %>%
  pivot_longer(cols = everything(), names_to = "Traitement", values_to = "Pourcentage")

ggplot(traitements, aes(x = Traitement, y = Pourcentage, fill = Traitement)) +
  geom_bar(stat = "identity", width = 0.7) +
  geom_text(aes(label = paste0(Pourcentage, "%")), vjust = -0.5, size = 4) +
  scale_fill_brewer(palette = "Set3") +
```

```

labs(title = "Traitements administrés",
     x = "", y = "Pourcentage (%)") +
theme_minimal() +
theme(legend.position = "none")

```

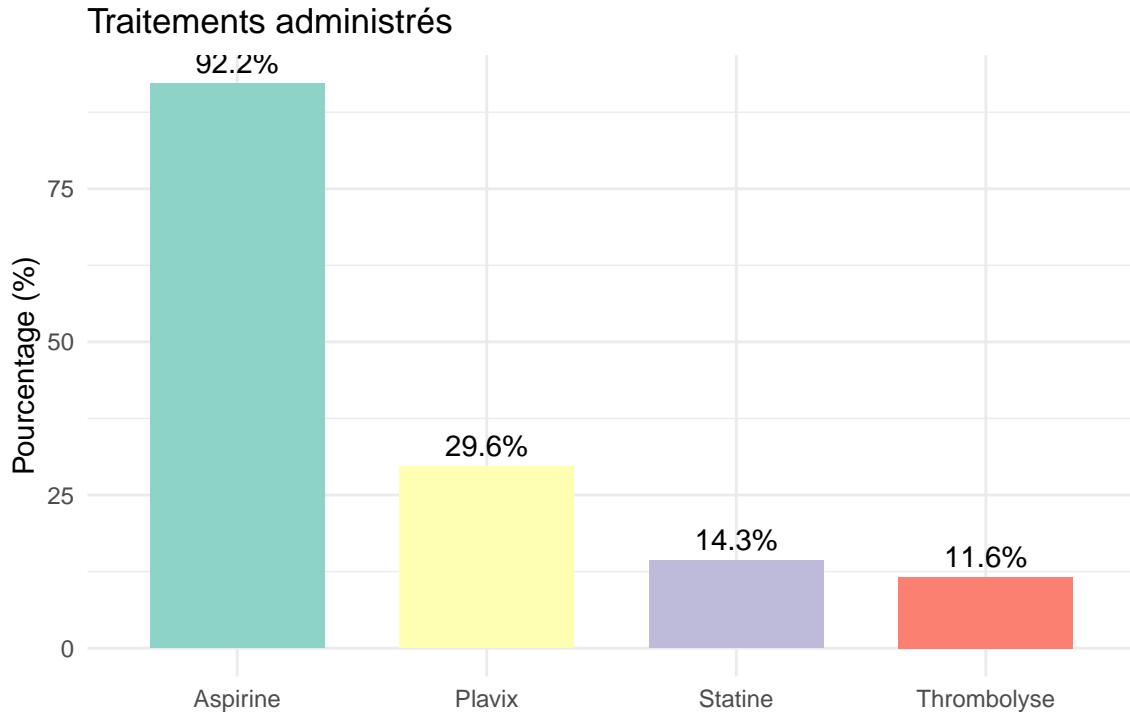


FIGURE 17 : Pourcentage de patients ayant reçu chaque traitement

## 6.2 Évolution du NIHSS dans le temps

```

nihss_temps <- avc_data %>%
summarise(
  Initial = round(mean(NIHSSinitial, na.rm = TRUE), 1),
  Sortie = round(mean(NIHSSsortie, na.rm = TRUE), 1),
  `1 mois` = round(mean(NIHSS.1mois, na.rm = TRUE), 1),
  `3 mois` = round(mean(NIHSS.3mois, na.rm = TRUE), 1)
) %>%
pivot_longer(cols = everything(), names_to = "Temps", values_to = "Score")

nihss_temps$Temps <- factor(nihss_temps$Temps,

```

```

levels = c("Initial", "Sortie", "1 mois", "3 mois"))

ggplot(nihss_temps, aes(x = Temps, y = Score, group = 1)) +
  geom_line(color = "#3498db", linewidth = 1.5) +
  geom_point(color = "#e74c3c", size = 4) +
  geom_text(aes(label = Score), vjust = -1.5, size = 5, fontface = "bold") +
  labs(title = "Évolution du score NIHSS moyen",
       subtitle = "Score plus bas = meilleure récupération",
       x = "", y = "Score NIHSS moyen") +
  theme_minimal()

```

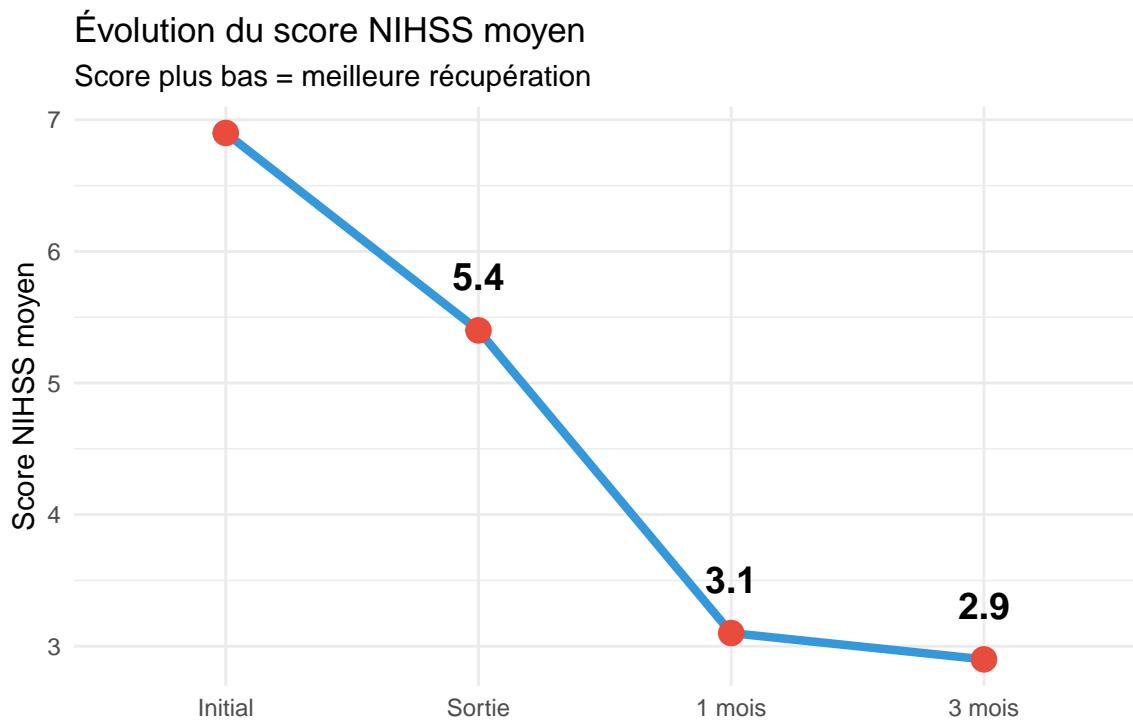


FIGURE 18 : Évolution du score NIHSS moyen à différents moments

### 6.3 Durée d'hospitalisation

```
cat("Moyenne:", round(mean(avc_data$Durée.Hospitalisation, na.rm = TRUE), 1), "jours\n")
```

Moyenne: 6.9 jours

```
cat("Médiane:", median(avc_data$Durée.Hospitalisation, na.rm = TRUE), "jours\n")
```

Médiane: 6 jours

```
ggplot(avc_data, aes(x = Durée.Hospitalisation)) +  
  geom_histogram(binwidth = 1, fill = "#2ecc71", color = "black", alpha = 0.8) +  
  geom_vline(aes(xintercept = mean(Durée.Hospitalisation, na.rm = TRUE)),  
             color = "darkgreen", linetype = "dashed", linewidth = 1) +  
  labs(title = "Distribution de la durée d'hospitalisation",  
        x = "Durée (jours)", y = "Nombre de patients") +  
  theme_minimal()
```

Distribution de la durée d'hospitalisation

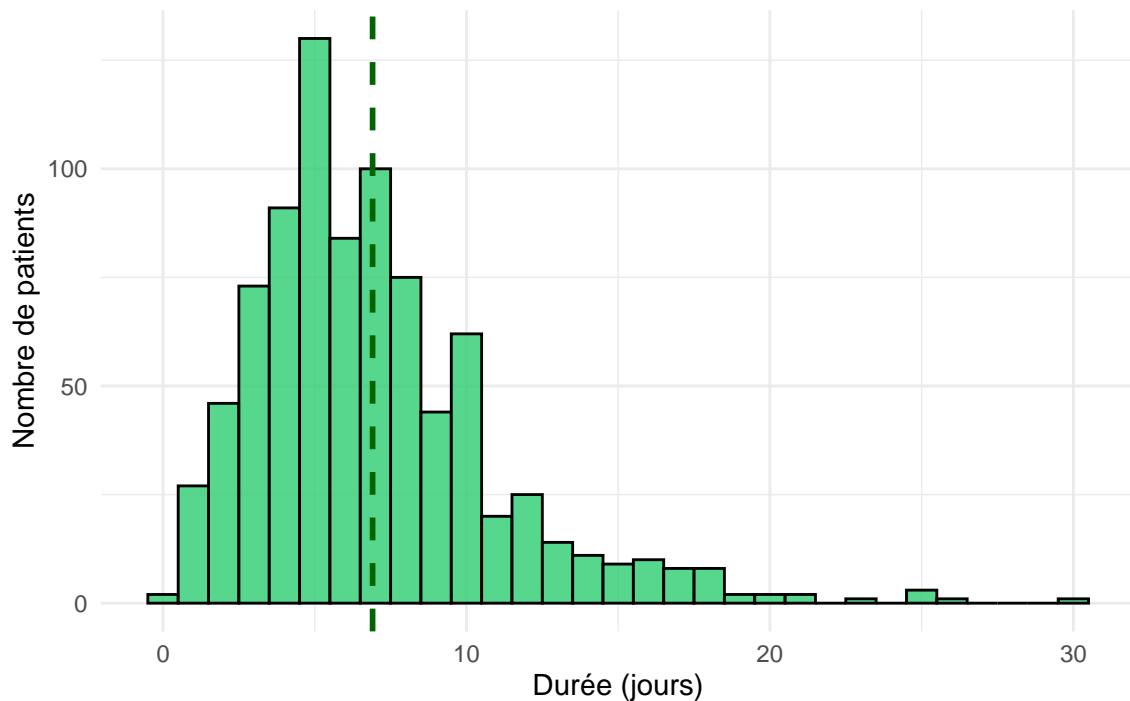


FIGURE 19 : Distribution de la durée d'hospitalisation

## 6.4 Évolution clinique

```
evolution_counts <- avc_data %>%  
  filter(!is.na(Evolution)) %>%  
  count(Evolution) %>%  
  mutate(
```

```

Pourcentage = n / sum(n) * 100,
Label = c("Amélioration", "Stable", "Aggravation")[Evolution]
)

ggplot(evolution_counts, aes(x = Label, y = n, fill = Label)) +
  geom_bar(stat = "identity", width = 0.6) +
  geom_text(aes(label = paste0(round(Pourcentage, 1), "%")),
            vjust = -0.5, size = 4) +
  scale_fill_manual(values = c("Amélioration" = "#2ecc71",
                               "Stable" = "#f1c40f",
                               "Aggravation" = "#e74c3c")) +
  labs(title = "Évolution clinique des patients",
       x = "", y = "Nombre de patients") +
  theme_minimal() +
  theme(legend.position = "none")

```

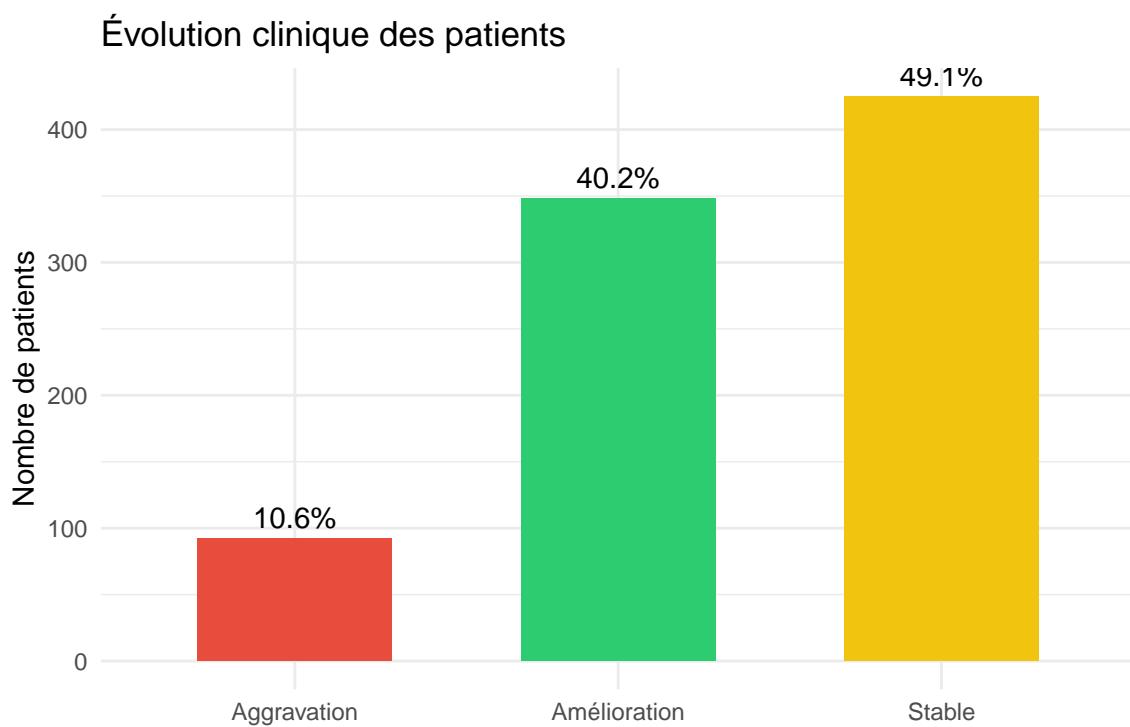


FIGURE 20 : Évolution clinique globale des patients

## 6.5 Taux de mortalité

```
deces_counts <- avc_data %>%
  filter(!is.na(Décés)) %>%
  count(Décés) %>%
  mutate(
    Pourcentage = n / sum(n) * 100,
    Décés = ifelse(Décés == 1, "Oui", "Non")
  )

ggplot(deces_counts, aes(x = Décés, y = n, fill = Décés)) +
  geom_bar(stat = "identity", width = 0.6) +
  geom_text(aes(label = paste0(round(Pourcentage, 1), "%")),
            vjust = -0.5, size = 5, fontface = "bold") +
  scale_fill_manual(values = c("Oui" = "#e74c3c", "Non" = "#2ecc71")) +
  labs(title = "Taux de mortalité",
       x = "Décès", y = "Nombre de patients") +
  theme_minimal() +
  theme(legend.position = "none")
```

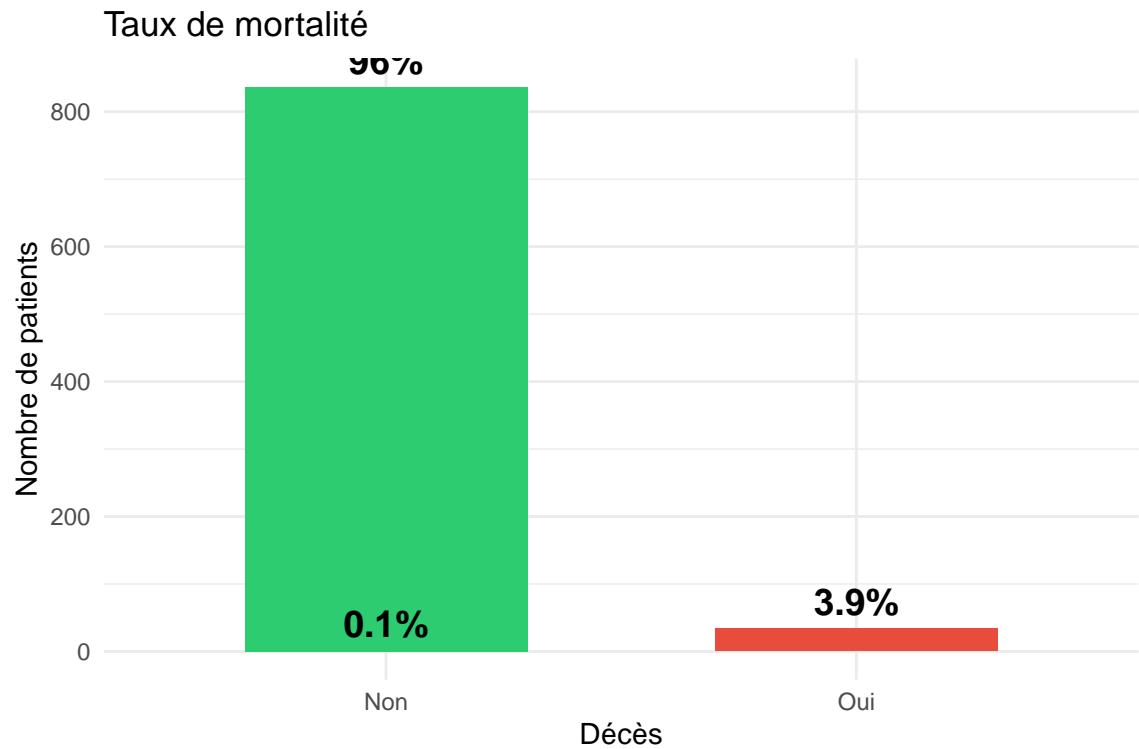


FIGURE 21 : Taux de mortalité dans la cohorte

## 6.6 Dashboard récapitulatif

```
p1 <- ggplot(sexe_counts, aes(x = Sexe, y = n, fill = Sexe)) +
  geom_bar(stat = "identity") +
  labs(title = "Répartition par sexe", x = "", y = "N") +
  theme_minimal() + theme(legend.position = "none")

p2 <- ggplot(facteurs, aes(x = Facteur, y = Pourcentage, fill = Facteur)) +
  geom_bar(stat = "identity") +
  coord_flip() +
  labs(title = "Facteurs de risque", x = "", y = "%") +
  theme_minimal() + theme(legend.position = "none")

p3 <- ggplot(nihss_temps, aes(x = Temps, y = Score, group = 1)) +
  geom_line(color = "#3498db", linewidth = 1) +
  geom_point(color = "#e74c3c", size = 2) +
  labs(title = "Évolution NIHSS", x = "", y = "Score") +
```

```

theme_minimal()

p4 <- ggplot(evolution_counts, aes(x = Label, y = n, fill = Label)) +
  geom_bar(stat = "identity") +
  labs(title = "Évolution clinique", x = "", y = "N") +
  theme_minimal() + theme(legend.position = "none")

grid.arrange(p1, p2, p3, p4, ncol = 2,
             top = "DASHBOARD RÉCAPITULATIF")

```

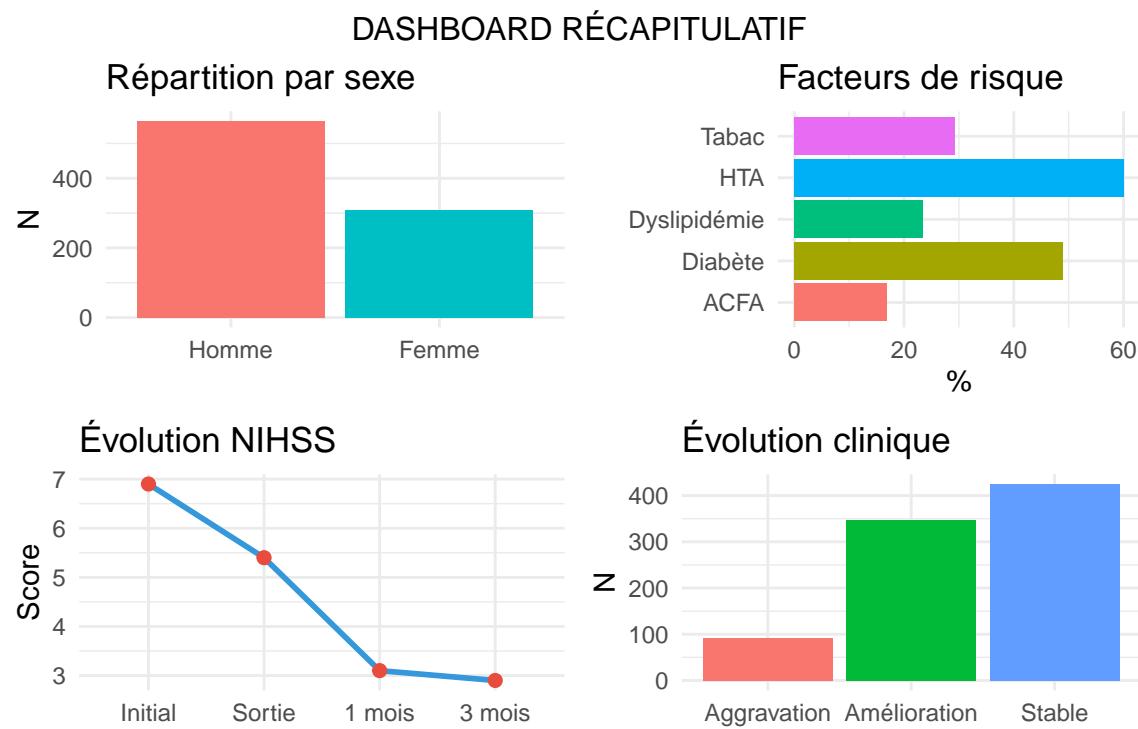


FIGURE 22 : Dashboard récapitulatif des principaux indicateurs

## 6.7 Analyse temporelle

```

# Graphique 1 : Histogramme des AVC par mois
avc_mois <- avc_data %>%
  mutate(
    DateAVC = as.Date(DateAVC),
    MoisNom = factor(month(DateAVC), levels = 1:12, labels = month.abb)

```

```

) %>%
filter(!is.na(MoisNom)) %>%
count(MoisNom)

p1_mois <- ggplot(avc_mois, aes(x = MoisNom, y = n)) +
  geom_bar(stat = "identity", fill = "#2ecc71", width = 0.7, alpha = 0.8) +
  geom_text(aes(label = n), vjust = -0.5, size = 4.5, fontface = "bold") +
  labs(
    title = "Distribution des AVC par mois",
    x = "Mois",
    y = "Nombre de cas"
  ) +
  theme_minimal(base_size = 12) +
  theme(
    axis.text.x = element_text(size = 11, angle = 45, hjust = 1),
    plot.title = element_text(size = 15, face = "bold", hjust = 0.5),
    panel.grid.major.x = element_blank()
  ) +
  scale_y_continuous(expand = expansion(mult = c(0, 0.1)))

# Graphique 2 : Histogramme Semaine vs Week-end
jours_counts <- avc_data %>%
  mutate(
    DateAVC = as.Date(DateAVC),
    NumJour = wday(DateAVC, week_start = 1),
    TypeJour = ifelse(NumJour %in% c(6, 7), "Week-end", "Semaine")
  ) %>%
  filter(!is.na(TypeJour)) %>%
  count(TypeJour) %>%
  mutate(Pourcentage = round(n / sum(n) * 100, 1))

p2_week <- ggplot(jours_counts, aes(x = TypeJour, y = n, fill = TypeJour)) +
  geom_col(width = 0.6) +
  geom_text(aes(label = paste0(n, " (", Pourcentage, "%)")), vjust = -0.5, fontface = "bold") +
  scale_fill_manual(values = c("Semaine" = "#3498db", "Week-end" = "#e74c3c")) +
  labs(title = "AVC : Semaine vs Week-end", x = "Période", y = "Nombre de cas") +
  theme_minimal() +

```

```

theme(legend.position = "none", plot.title = element_text(size = 15, face = "bold", hjust = 0.5)

# Affichage des deux graphiques côte à côté
grid.arrange(p1_mois, p2_week, ncol = 2)

```

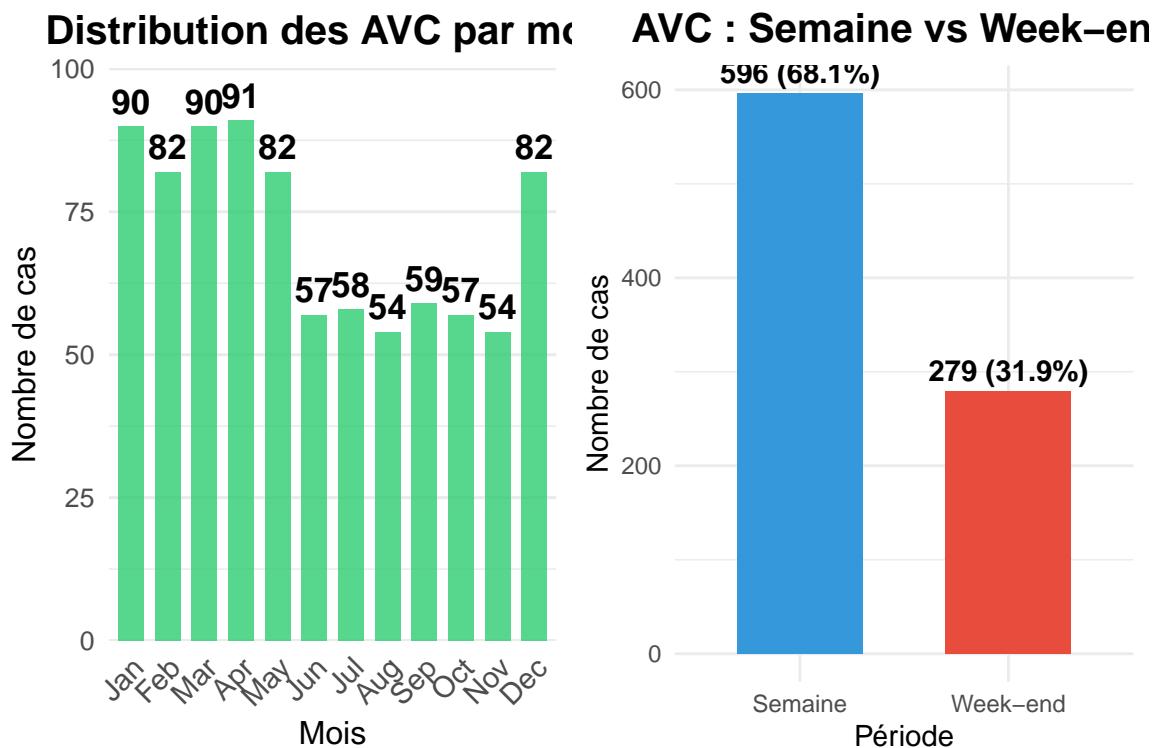


FIGURE 23 : Analyse temporelle : distribution par mois et semaine vs week-end

```

avc_data <- azt %>%
  mutate(
    DateAVC = as.Date(DateAVC),
    NumJour = wday(DateAVC, week_start = 1),
    JourSemaine = factor(NumJour,
      levels = 1:7,
      labels = c("Lundi", "Mardi", "Mercredi", "Jeudi", "Vendredi", "Samedi",
    )
  )

jours_counts <- azt %>%
  filter(!is.na(JourSemaine)) %>%
  count(JourSemaine) %>%

```

```

mutate(
  Pourcentage = round(n / sum(n) * 100, 1),
  TypeJour = ifelse(JourSemaine %in% c("Samedi", "Dimanche"), "Week-end", "Semaine")
)

ggplot(jours_counts, aes(x = JourSemaine, y = n, fill = TypeJour)) +
  geom_col(width = 0.7) +
  geom_text(aes(label = paste0(n, "\n(", Pourcentage, "%)")),
            vjust = 1.2, color = "white", fontface = "bold") +
  scale_fill_manual(values = c("Semaine" = "#3498db", "Week-end" = "#e74c3c")) +
  theme_minimal() +
  labs(title = "Distribution des AVC par jour de la semaine", x = "Jour", y = "Nombre")

```

Distribution des AVC par jour de la semaine

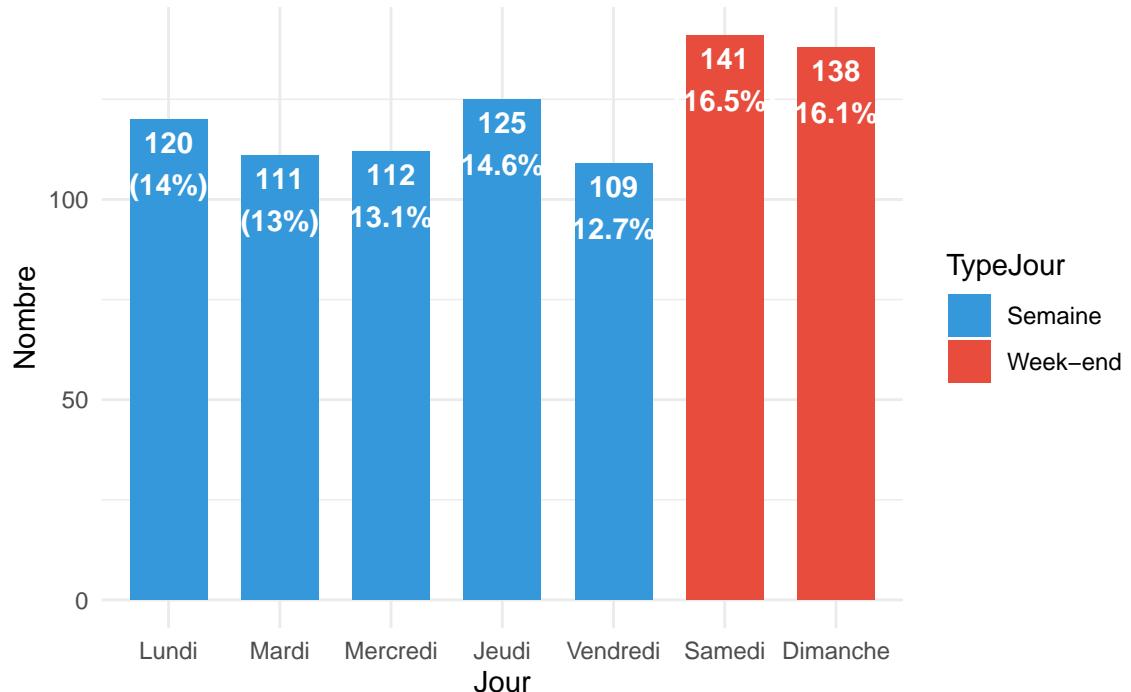


FIGURE 24 : Distribution des AVC par jour de la semaine

## 7 Conclusions

### 7.1 Rapport statistique final

#### 7.1.1 1. Caractéristiques de l'échantillon

- **Nombre total :** 875 patients
- **Âge moyen :**  $64.3 \pm 13$  ans
- **Sexe ratio H/F :** 1.83

#### 7.1.2 2. Facteurs de risque principaux

- **HTA :** 60.1%
- **Diabète :** 49%
- **Tabac :** 29.2%

#### 7.1.3 3. Sévérité et évolution

- **NIHSS initial moyen :** 6.85
- **NIHSS sortie moyen :** 5.43
- **Amélioration moyenne :** 1.35 points

#### 7.1.4 4. Prise en charge

- **Thrombolyse :** 11.6%
- **Durée d'hospitalisation :** 6.9 jours
- **Taux de mortalité :** 3.9%

#### 7.1.5 5. Tests statistiques principaux

- **Test t âge H/F :**  $p = 2.92\text{e-}06$
- **Test t apparié NIHSS :**  $p = 4.19\text{e-}21$
- **Corrélation Âge-NIHSS :**  $r = 0.179$

## 7.2 Résumé des principales découvertes

### 1. Caractéristiques de la population :

- Population principalement âgée (moyenne 64.3 ans)
- Prédominance masculine (64.7% hommes vs 35.3% femmes)
- Les femmes ont des AVC significativement plus tard (67 ans vs 63 ans)

### 2. Facteurs de risque :

- HTA : présent chez 60% des patients (facteur de risque majeur)

- Diabète : présent chez 49% des patients
- Tabac : 29% de fumeurs
- Ces facteurs de risque modifiables sont très fréquents

### **3. Sévérité et pronostic :**

- Score NIHSS moyen initial : 6.85 (sévérité légère à modérée)
- Amélioration significative pendant l'hospitalisation (NIHSS 6.78 → 5.44)
- Récupération continue après la sortie (amélioration à 1 et 3 mois)
- Taux de mortalité faible : 3.9%

### **4. Prise en charge :**

- Délai de consultation excellent (médiane : 6 minutes)
- Thrombolyse administrée à 11.6% des patients éligibles
- Aspirine administrée à 92% des patients (traitement standard)
- Durée d'hospitalisation moyenne : 6.9 jours

### **5. Évolution clinique :**

- 40% des patients s'améliorent
- 49% restent stables
- 10% s'aggravent
- 90% ont une évolution favorable ou stable

## **7.3 Implications cliniques**

- 1. Prévention :** L'HTA et le diabète sont les facteurs de risque les plus fréquents. Des programmes de prévention ciblant ces facteurs pourraient réduire significativement l'incidence des AVC.
- 2. Prise en charge :** Le délai de consultation rapide permet un traitement optimal. La majorité des patients répondent bien au traitement standard.
- 3. Facteurs de risque modifiables :** Le tabac, l'HTA et le diabète sont des facteurs modifiables. Des interventions sur ces facteurs pourraient améliorer les résultats.
- 4. Différences selon le sexe :** Les femmes ont des AVC plus tard, peut-être liés à des facteurs hormonaux. Des stratégies de prévention différencieront pourraient être bénéfiques.

## **7.4 Limitations**

1. Données manquantes dans certaines variables (notamment analyses biologiques)
2. Données provenant d'un seul centre (généralisation limitée)
3. Pas d'analyse multivariée approfondie dans cette analyse exploratoire
4. Certaines variables ont peu de données

## **7.5 Recommandations**

1. Améliorer le contrôle de l'HTA et du diabète dans la population à risque
2. Promouvoir l'arrêt du tabac
3. Maintenir les délais de consultation rapides
4. Assurer un suivi régulier post-AVC pour optimiser la récupération
5. Envisager des analyses multivariées pour identifier les facteurs de risque indépendants