

# DIABETIES DATASET

---

L3MIAGE

---

Yassmine TAI TAHIRI, Hadil CHTIOUI,  
Omar El Farouk CHTIOUI, Nada BOUSSETTA

Projet Modèles Statistiques

# SOMMAIRE

1. Introduction
2. Objectif principal
3. Méthodologie
4. Analyse
5. Conclusion





# INTRODUCTION



kaggle

100 000 individus

## OBJECTIF PRINCIPAL



**“Quels sont les facteurs les plus associés à la présence de diabète chez un individu ? ”**

# MÉTHODOLOGIE



## Structure des données

- dim()
- colnames()
- summary()
- str()



## Valeurs possibles pour chaque colonnes

- Pour comprendre le type de chaque colonne et les valeurs qu'elles peuvent prendre



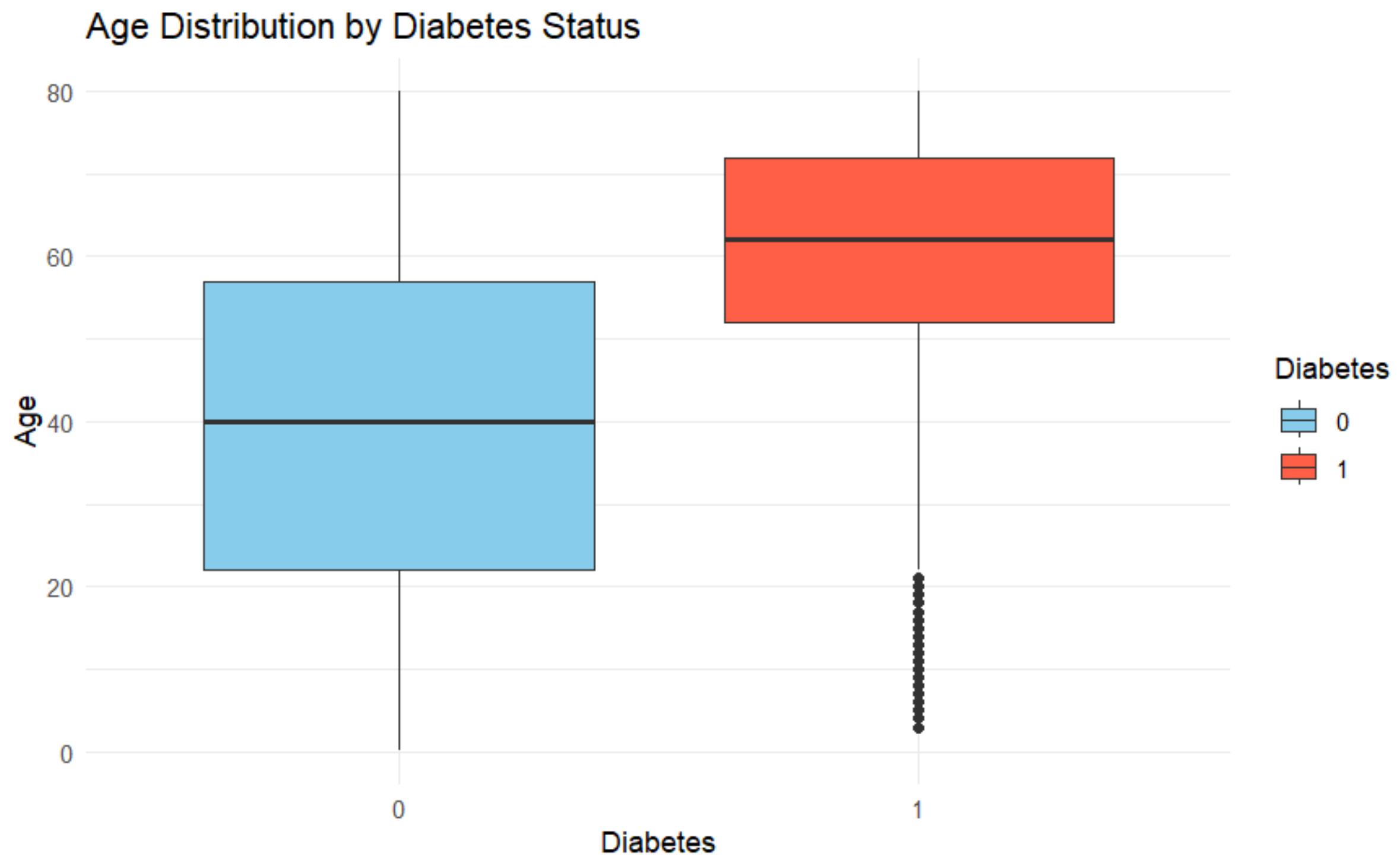
## Nettoyage des données

- Supprimer les doublons et les valeurs manquantes
- Supprimer les observations non significatives ( comme 'no info' pour le smoking\_history)

# ANALYSE

## Analyse univariée

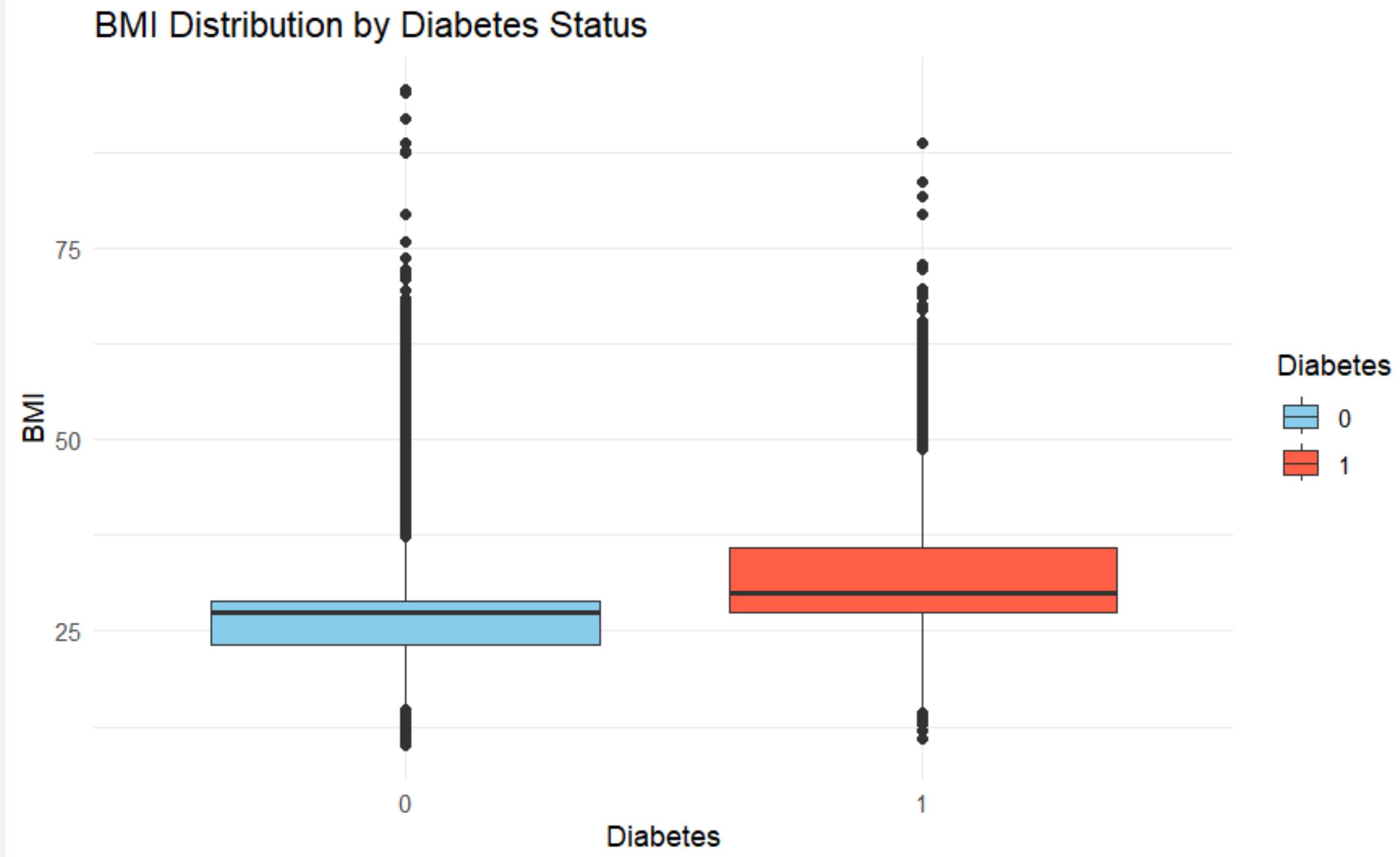
Exemple 1 : L'âge



# ANALYSE

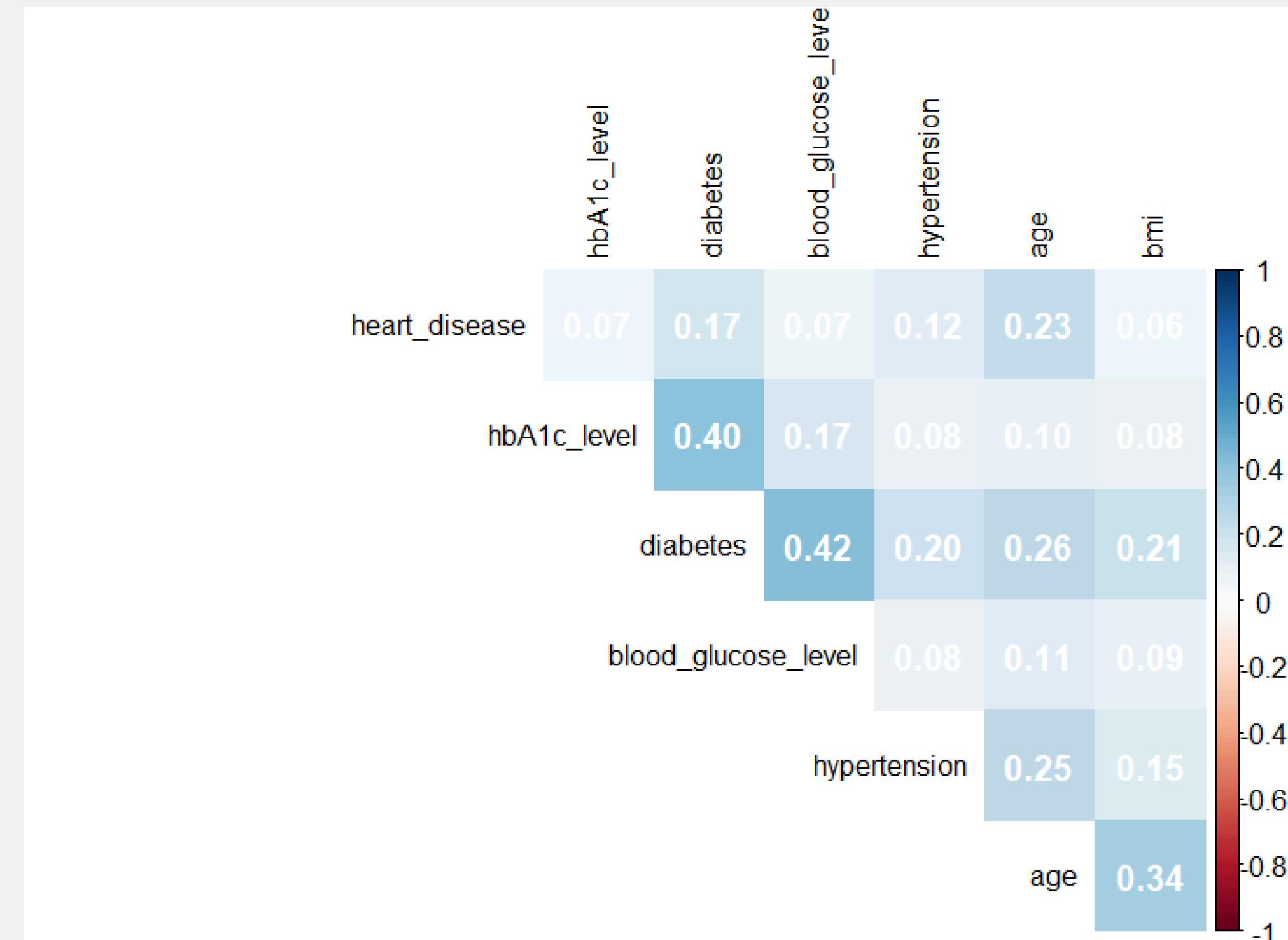
## Analyse univariée

Exemple 2 : BMI



# ANALYSE

## Corrélation



# ANALYSE

## Analyse croisée

