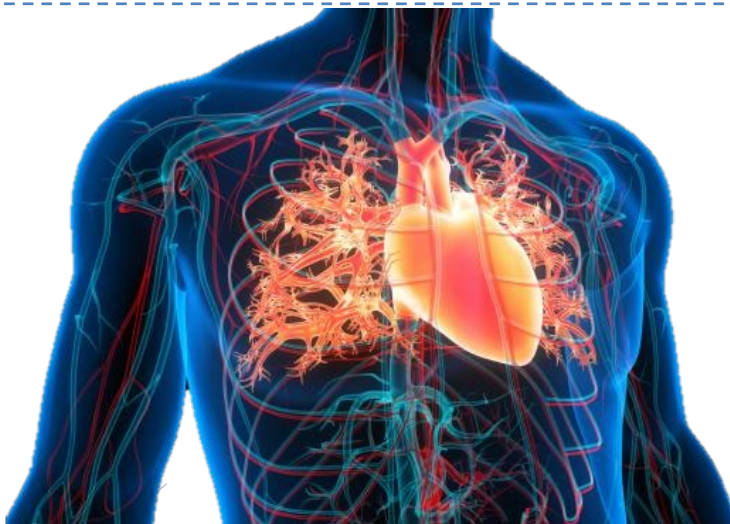


30/11/2023

**ANNEE UNIVERSITAIRE:
2023-2024**

MINI-PROJET 1



THEME

**ANALYSE DES FACTEURS
CONTRIBUANT AU DIAGNOSTIQUE
DES MALADIES CARDIAQUES**

Encadré par :

M. Akposso Didier Martial

Réalisé par :

M. Savané Mory

SOMMAIRE

INTRODUCTION GENERALE.....	0
PREMIERE PARTIE : PRETRAITEMENT DES DONNEES	3
Introduction.....	3
I. DESCRIPTION DES DONNEES.....	3
II. TRAITEMENT DES VALEURS MANQUANTES, VALEURS ABERRANTES ET EXTREMES.....	7
Conclusion	10
DEUXIEME PARTIE : ANALYSE UNIVARIEE.....	11
Introduction.....	11
I. ANALYSE DES VARIABLES QUANTITATIVES.....	11
II. ANALYSE DES VARIABLES QUALITATIVES.....	16
Conclusion	24
TROISIEME PARTIE : ANALYSE BIVARIEE.....	25
Introduction.....	25
II. Liaison entre Les maladies cardiaques et les variables qualitatives.....	32
1. Les maladies cardiaques sont-elles influencées par le genre ?.....	32
2. <i>Les maladies cardiaques sont-elles influencées par l'Angine de poitrine ?</i>	35
3. Les maladies cardiaques sont-elles influencées par la glycémie à jeun ?	38
4. Les maladies cardiaques sont-elles influencées par le résultat de l'électrocardiogramme ?	41
5. Les maladies cardiaques sont-elles influencées par la pente de l'ECG ?.....	44
6. <i>Les maladies cardiaques sont-elles influencées par l'Angine causée par l'exercice ?</i>	47
Conclusion	49
CONCLUSION GENERALE	38

INTRODUCTION GENERALE

Les maladies cardiaques demeurent un enjeu majeur de santé publique, représentant une cause significative de morbidité et de mortalité à l'échelle mondiale.

Selon l'OMS¹, il meurt chaque année plus de personnes en raison de maladies cardio-vasculaires que de toute autre cause. On estime à 17,7 millions le nombre de décès imputables aux maladies cardio-vasculaires, soit 31% de la mortalité mondiale totale. Parmi ces décès, on estime que 7,4 millions sont dus à une cardiopathie coronarienne et 6,7 millions à un AVC (chiffres 2015). (OMS 2017)

Comprendre les facteurs qui contribuent au diagnostic précis de ces affections revêt une importance capitale pour améliorer les stratégies de prévention et de prise en charge.

Alors face à la complexité des maladies cardiaques, il devient impératif d'identifier des marqueurs fiables et des indicateurs précis pour un diagnostic précoce. La question centrale qui guide notre démarche est la suivante : **quels sont les facteurs déterminants qui contribuent de manière significative au diagnostic des maladies cardiaques ?**

Notre étude se propose donc d'explorer en profondeur ces aspects, en mettant en lumière des éclairages significatifs obtenus à partir d'une analyse approfondie des données.

Pour ce faire, notre démarche s'articulera autour de trois grandes parties. Tout d'abord, nous aborderons *le prétraitement des données*, où nous mettrons en œuvre des méthodes rigoureuses pour garantir la qualité et la cohérence des informations recueillies. Ensuite, nous procéderons à une *analyse univariée*, examinant chaque variable de manière isolée pour déceler des tendances ou des relations potentielles avec les maladies cardiaques. Enfin, la troisième partie portera *sur l'analyse bivariée*, permettant d'explorer les interactions entre différentes variables et d'identifier des associations significatives.

Cette approche structurée nous permettra d'obtenir des éclairages pertinents et d'enrichir la base de connaissances sur les maladies cardiaques, contribuant ainsi à l'amélioration des stratégies de prévention et de prise en charge.

¹ Organisation Mondiale de la Santé

PREMIERE PARTIE : PRETRAITEMENT DES DONNEES

Introduction

Selon une étude menée par IBM, la mauvaise qualité des données coûte désormais 3 100 milliards de dollars par an aux États-Unis. Et ce coût augmente de façon exponentielle. Cette étude met en exergue l'importance d'avoir des données de très bonnes qualités afin d'avoir des résultats pertinents et fiables. Le nettoyage des données ou data cleansing est donc une étape essentielle de l'analyse statistique et de la Data Science. Il s'agit de corriger ou supprimer des enregistrements inexacts dans des jeux de données afin pouvoir les exploiter par la suite. C'est un processus qui vise, de manière générale, à améliorer **la qualité des données**. En pratique, il consiste à **importer les données** dans un logiciel statistique, dans notre cas l'IDE Rstudio du logiciel R. Ensuite, faire une **première visualisation** pour avoir une idée succincte de la structure de nos données et enfin, identifier et corriger les valeurs manquantes, valeurs aberrantes, valeurs extrêmes afin de qu'elles puissent être plus cohérentes et sans erreurs.

I. DESCRIPTION DES DONNEES

1. Dictionnaire des données

Le dictionnaire de donnée ou spécification fonctionnelle des données est un document qui renseigne sur le contexte d'une base de données et qui fournit les informations nécessaires pour pouvoir l'interpréter. C'est donc un outil très important qui facilite la compréhension de la structure des données pour les administrateurs et les utilisateurs. Dans le cadre de notre étude ce document se présente comme suit :

Tableau 1: Dictionnaire des données

VARIABLE	NATURE	DESCRIPTION	MODALITES
Age	Quantitative	âge du patient	Numérique Entier en années
Sex	Qualitative	sexe/genre du patient	M : Masculin ou F : Feminin
ChestPainType	Qualitative	Type de douleur thoracique	TA : Angine de poitrine typique, ATA : Angine de poitrine atypique, NAP: Douleur autre que l'angine,

			ASY: Asymptomatique
RestingBP	Quantitative	Tension artérielle au repos (en mm Hg)	Numerique Decimal
Cholesterol	Quantitative	Taux de cholestérol (en mm/dl)	Numerique Decimal
FastingBS	Qualitative	glycémie à jeun	Qualitative Binaire (1 si FastingBS > 120 mg/dl, 0 sinon)
RestingECG	Qualitative	électrocardiogramme au repos	LVH, Normal et ST
MaxHR	Quantitative	Fréquence cardiaque maximale (en battements par minutes)	Numerique Decimal
ExerciseAngina	Qualitative	angine induite par l'exercice	Non.angine (Présence d'angine apres exercice) et Oui.angine(Absence d'angine apres exercice)
Oldpeak	Quantitative	Dépression du segment ST induite par l'exercice	Numerique Decimal
ST_Slope	Qualitative	pente du segment ST	Up: croissante ; Flat : constante ; Down : décroissante
HeartDisease	Qualitative	Présence ou absence de maladie cardiaque	Non ou Oui

2. Visualisation du jeu de donnée brute et de sa structure

Avant de débiter tout analyse, il est primordial de faire une première visualisation des données dans le but de se faire une idée succincte de la répartition des données, des noms des différentes colonnes et de l'agencement des différentes lignes. Par ailleurs, cette première phase est appuyée par la visualisation de la structure et de la statistique descriptive du dataframe.

Tableau 2: Aperçu du jeu de données

Age	Sex	ChestPainType	RestingBP	Cholesterol	FastingBS	RestingECG	MaxHR	ExerciseAngina	Oldpeak	ST_Slope	HeartDisease
40	M	ATA	140	289	0	Normal	172	N	0.0	Up	0
49	F	NAP	160	180	0	Normal	156	N	1.0	Flat	1
37	M	ATA	130	283	0	ST	98	N	0.0	Up	0
48	F	ASY	138	214	0	Normal	108	Y	1.5	Flat	1
54	M	NAP	150	195	0	Normal	122	N	0.0	Up	0

(Source: Kaggle, nos calculs 2021)

Ce premier aperçu met à nu *quelques anomalies* au niveau des types des variables **FastingBS** et **HeartDisease** qui sont considérées comme des variables quantitatives alors qu'elles sont qualitatives. Nous allons donc les recoder en variables qualitatives.

Tableau 3: Structure avant recodage

'data.frame': 918 obs. of 12 variables:											
\$ Age	:	int	40 49 37 48 54 39 45 54 37 48 ...								
\$ Sex	:	Factor w/ 2 levels "F","M":	2 1 2 1 2 2 1 2 2 1 ...								
\$ ChestPainType	:	Factor w/ 4 levels "ASY","ATA","NAP",...:	2 3 2 1 3 3 2 2 1 2 ...								
\$ RestingBP	:	int	140 160 130 138 150 120 130 110 140 120 ...								
\$ Cholesterol	:	int	289 180 283 214 195 339 237 208 207 284 ...								
\$ FastingBS	:	int	0 0 0 0 0 0 0 0 0 0 ...								
\$ RestingECG	:	Factor w/ 3 levels "LVH","Normal",...:	2 2 3 2 2 2 2 2 2 2 ...								
\$ MaxHR	:	int	172 156 98 108 122 170 170 142 130 120 ...								
\$ ExerciseAngina	:	Factor w/ 2 levels "N","Y":	1 1 1 2 1 1 1 1 2 1 ...								
\$ Oldpeak	:	num	0 1 0 1.5 0 0 0 0 1.5 0 ...								
\$ ST_Slope	:	Factor w/ 3 levels "Down","Flat",...:	3 2 3 2 3 3 3 3 2 3 ...								
\$ HeartDisease	:	int	0 1 0 1 0 0 0 1 0 ...								

Recodage de certaines variables

Suite au recodage, nous pouvons aisément remarquer des changements au niveau des variables : Sex; FastingBS; ExerciseAngina et HeartDisease

Tableau 4: Structure post recodage

'data.frame': 918 obs. of 12 variables:											
\$ Age	:	int	40	49	37	48	54	39	45	54	37 48 ...
\$ Sex	:	Factor w/ 2 levels	"Femme",	"Homme":	2	1	2	1	2	2	1 2 2 1 ...
\$ ChestPainType	:	Factor w/ 4 levels	"ASY",	"ATA",	"NAP",	..:	2	3	2	1	3 3 2 2 1 2 ...
\$ RestingBP	:	int	140	160	130	138	150	120	130	110	140 120 ...
\$ Cholesterol	:	int	289	180	283	214	195	339	237	208	207 284 ...
\$ FastingBS	:	Factor w/ 2 levels	"FastingBS < 120 mg/dl",	..:	1	1	1	1	1	1	1 1 ...
\$ RestingECG	:	Factor w/ 3 levels	"LVH",	"Normal",	..:	2	2	3	2	2	2 2 2 2 2 ...
\$ MaxHR	:	int	172	156	98	108	122	170	170	142	130 120 ...
\$ ExerciseAngina	:	Factor w/ 2 levels	"N",	"Y":	1	1	1	2	1	1	1 1 2 1 ...
\$ Oldpeak	:	num	0	1	0	1.5	0	0	0	0	1.5 0 ...
\$ ST_Slope	:	Factor w/ 3 levels	"Down",	"Flat",	..:	3	2	3	2	3	3 3 3 2 3 ...
\$ HeartDisease	:	Factor w/ 2 levels	"Non",	"Oui":	1	2	1	2	1	1	1 1 2 1 ...

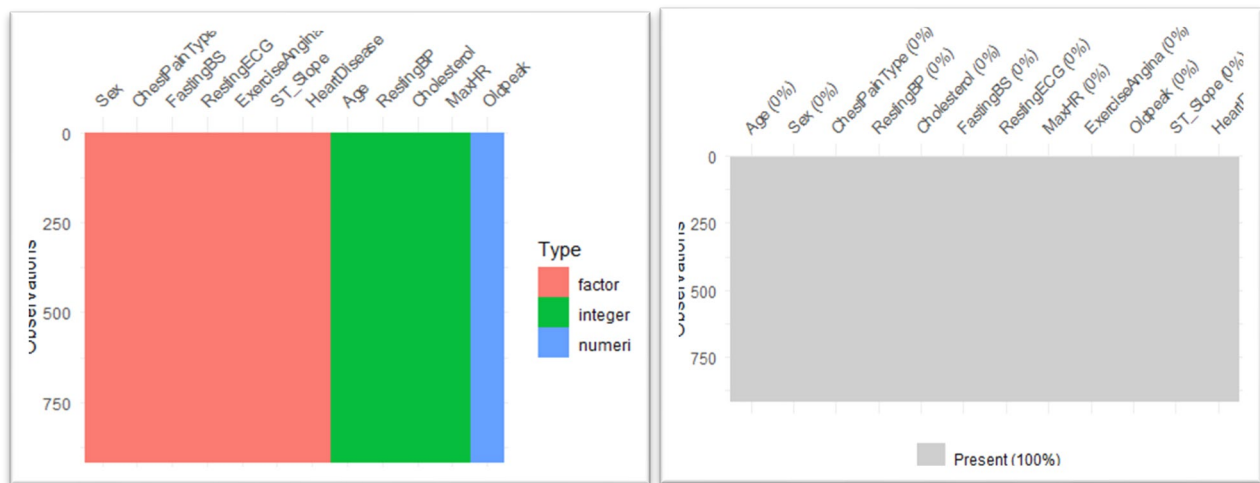
Cet aperçu nous permet de savoir que nous sommes en présence de **918 observations** avec **douze (12) variables** dont **5 quantitatives** et **7 qualitatives**.

II. TRAITEMENT DES VALEURS MANQUANTES, VALEURS ABERRANTES ET EXTREMES

Une valeur manquante est une donnée qui n'est pas présente pour une variable du jeu de données soumis à notre analyse. Il est généralement matérialisé dans les logiciels d'analyse par NA (Not Available).

1. Visualisation des données manquantes

Figure 1: Visualisation des Valeurs manquantes



Suite à la visualisation, nous remarquons que nos données sont exemptes de valeurs manquantes, ce qui nous donne une avance considérable dans la suite de notre analyse.

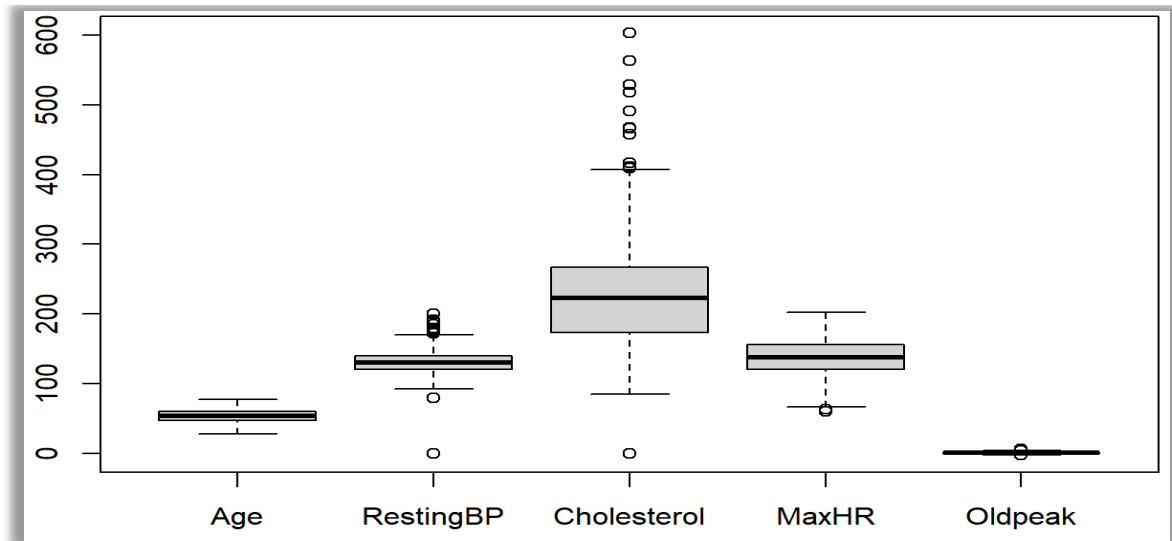
2. Visualisation des valeurs aberrantes et extrêmes

Nous débutons par extraire les variables quantitatives et ensuite, nous visualisons les boîtes à moustaches associées.

Tableau 5: Aperçu des variables quantitatives

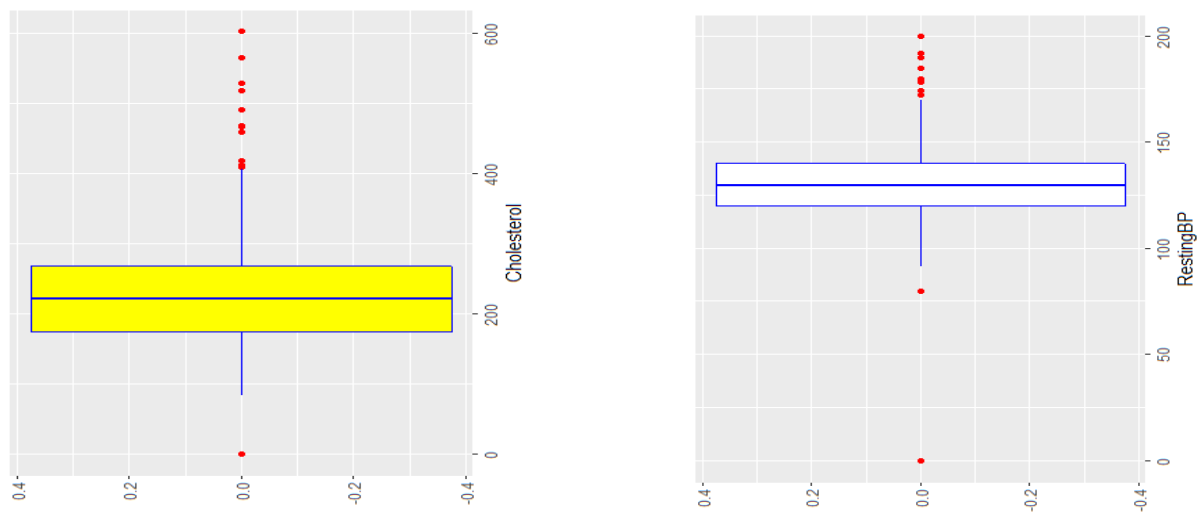
Age	RestingBP	Cholesterol	MaxHR	Oldpeak
40	140	289	172	0.0
49	160	180	156	1.0
37	130	283	98	0.0
48	138	214	108	1.5
54	150	195	122	0.0
39	120	339	170	0.0

Figure 2: Boîte à Moustache des variables quantitatives



La visualisation des boites a moustache des variables quantitatives nous permet de focaliser notre attention sur deux principales variables : la variable **RestingBP** et la variable **Cholesterol**, car elles présentent des valeurs aberrantes.

a. Visualisation des valeurs aberrantes et extrêmes



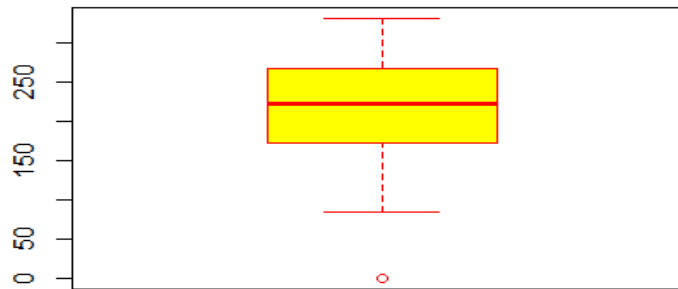
La visualisation approfondie des boites à moustache de ces deux variables **confirme la présence de valeur aberrante et extrême**.

Nous allons donc utiliser **la méthode de winzorisation** pour ramener ces valeurs aux différentes bornes inférieurs et supérieurs.

b. Traitement des valeurs aberrantes et extrêmes

- *Identification et traitement des valeurs aberrantes de la variable **Cholestérol***

CHOLESTEROL APRES WINZORISATION

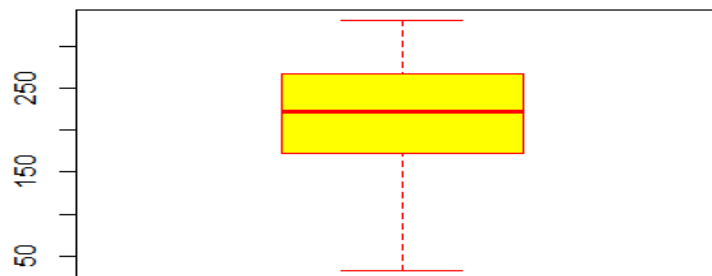


La méthode de Winzorisation a traité les valeurs aberrantes de la moustache supérieurs, **mais a maintenu les individus ayant des valeurs égales à 0**, qui sont effectivement des valeurs aberrantes. Nous leur attribuerons la valeur de la moustache inférieure. Qui a la valeur suivante :

```
## [1] 32.625 32.625 32.625 32.625 32.625 32.625
```

La boîte a moustache ci-dessus prouve que toutes les valeurs aberrantes et extrêmes ont été traitées.

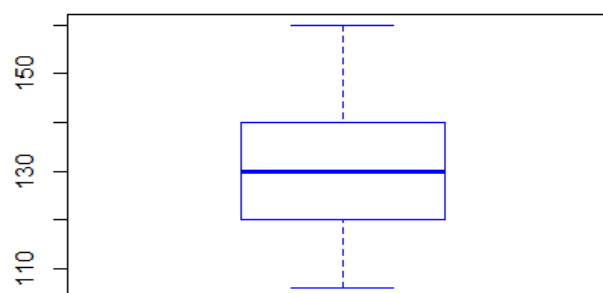
CHOLESTEROL APRES ATTRIBUTION



- *Identification et traitement des valeurs aberrantes de la variable **RestingBP***

Cette fois, l'application de la méthode de Winzorisation a permis de traiter toutes les valeurs aberrantes et extrêmes. Nous obtenons le résultat suivant :

TENSION ARTERIELLE APRES WINZORISATION



Conclusion

La phase de prétraitement de nos données nous a permis de les débarrasser de tout ce qui pourrait mettre en question la pertinence des résultats obtenus à la suite de notre analyse.

DEUXIEME PARTIE : ANALYSE UNIVARIEE

Introduction

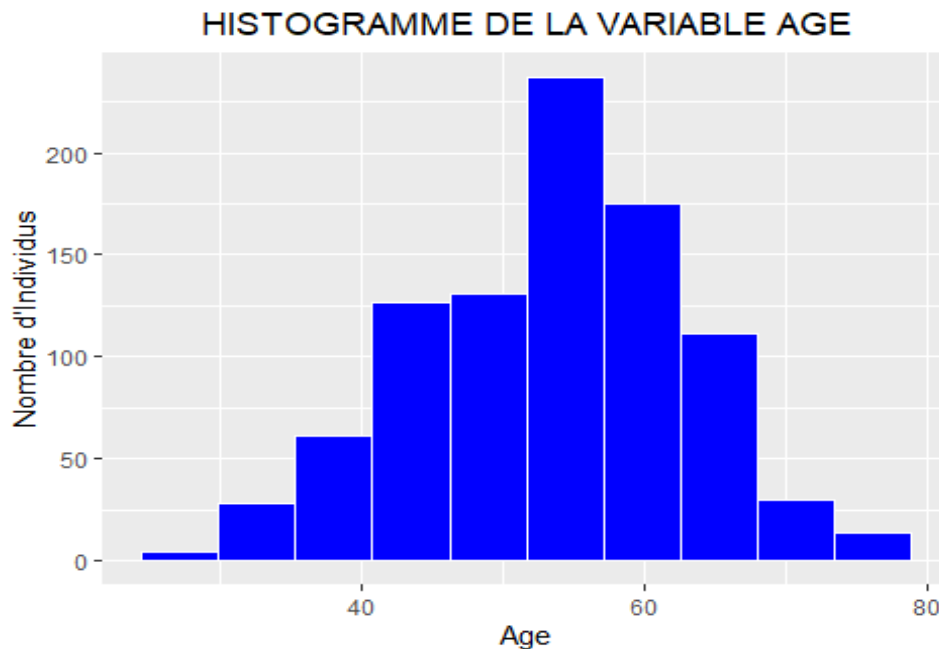
L'Analyse univariée consiste à visualiser la distribution de nos données grâce à des graphiques adaptés à chaque type de variable, à calculer et interpréter les indicateurs de tendance centrale, les indicateurs de position et de forme.

I. ANALYSE DES VARIABLES QUANTITATIVES

Pour chaque variable, l'analyse se fera en trois parties : **la représentation des données sous forme de graphique, sous forme de tableau numérique.**

Par ailleurs, chaque partie sera accompagnée d'une interprétation.

1. Analyse et interprétation de l'âge des individus



La visualisation de l'histogramme nous donne un aperçu visuel de la répartition des données. L'âge des individus de notre population est compris entre **28 ans et 77 ans**. Ils ont un **âge moyen** s'élevant à **54 ans et demi**, avec un écart type d'environ 9 ans.

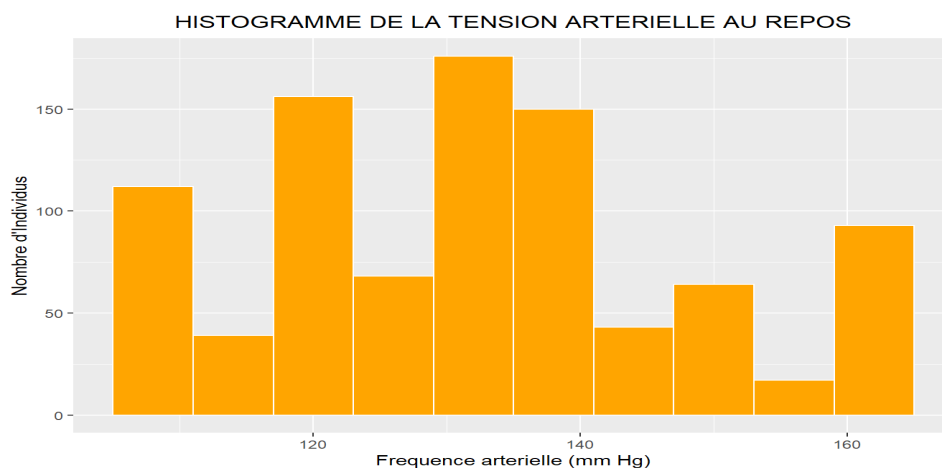
De plus, **un individu sur quatre a au moins de 47 ans** tandis que **l'âge médian s'élève à 54 ans**, c'est à dire qu'un individu sur deux a 54 ans. Cela dénote d'une **population majoritairement adulte**.

2. Analyse et interprétation de la tension artérielle des individus (RestingBP)

La pression artérielle au repos représente la pression exercée par le sang sur les parois des artères lorsque le corps est au repos, en particulier *entre les battements cardiaques*.

Sa mesure se fait à l'aide d'un tensiomètre et est généralement prise lorsque la personne est assise ou allongée de manière confortable, et elle donne une indication de la pression sanguine de base dans le système circulatoire.

Une pression artérielle au repos *normale* est généralement autour de *120/80 mm Hg* (Millimètre de Mercure) où 80 mm Hg est la valeur de la fréquence artérielle au repos. La visualisation de



L'histogramme nous donne un aperçu visuel de la répartition de la tension artérielle.

La tension des individus de notre population est comprise entre 106 mm Hg et 160 mm Hg, alors que la **valeur modale est de 120 mm Hg**. Cela met en exergue que la plupart des personnes ont une *tension artérielle qui est autour de la normale*.

La fréquence artérielle moyenne s'élève à 132 mm Hg, avec un écart type d'environ 16 mm Hg. De plus, un individu sur quatre a au plus 120 mm Hg tandis que la pression artérielle médiane est de 130 mm Hg, c'est à dire qu'**un individu sur deux a 130 mm Hg**, ce qui est relativement au-dessus de la norme qui est 120 mm Hg.

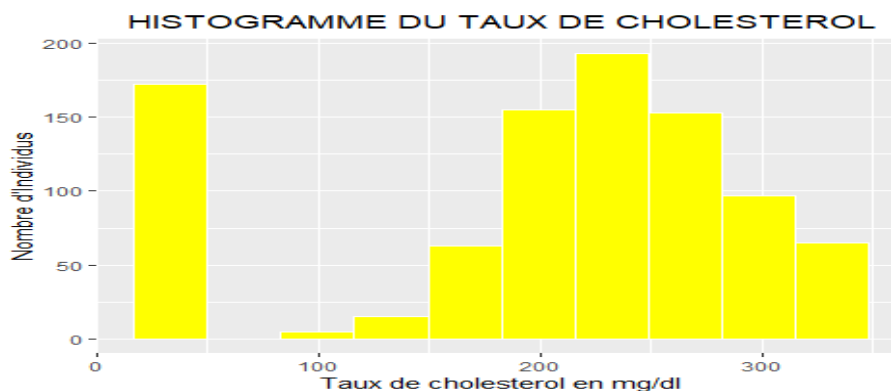
3. Analyse et interprétation du taux de cholestérol des individus

L'organisme produit près de 75% du cholestérol. Le reste provient de l'alimentation. On le retrouve dans les produits d'origine animale (viande, charcuteries, laitages, abats...) ainsi que dans les œufs, poissons, crustacés. Il est synthétisé en grande partie par le foie. **La régulation du cholestérol dépend donc de facteurs génétiques et de notre mode d'alimentation.**

Il existe principalement deux types de cholestérol, le **bon cholestérol (HDL)** ou High Density Lipoprotein et le **mauvais cholestérol (LDL)** ou Low Density Lipoprotein.

Un équilibre approprié entre le LDL et le HDL est important pour maintenir la santé cardiovasculaire. Des niveaux élevés de LDL et des niveaux bas de HDL peuvent augmenter le risque de maladies cardiovasculaires.

Le Cholestérol Total (CT) est la combinaison des deux sus-cites. Le taux de cholestérol total ne devrait pas être supérieur à 200, voire 238 mg/dl. Lorsque ce chiffre dépasse les 240 mg/dl, cela peut



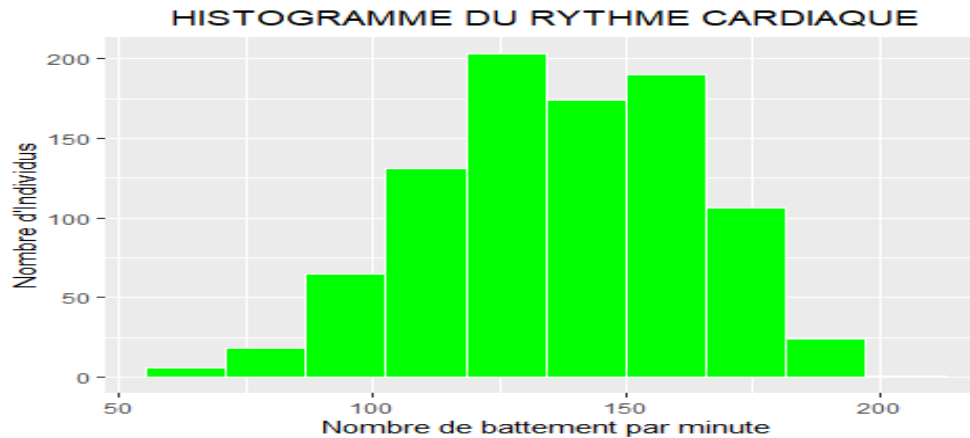
même s'avérer dangereux.

Le niveau du taux de cholestérol des individus de notre population est compris entre 32.62 mg/dl et 331.3 mg/dl, alors que la valeur modale est de 222.56 mg/dl. Cela met en exergue que **la plupart des personnes ont un taux de cholestérol autour de la valeur normale qui est de 200 mg/dl.** Cependant, ce taux peut parfois atteindre 300 mg/dl, ce qui pourrait s'avérer très dangereux.

Le **taux moyen de cholestérol s'élève à 202 mg/dl**, avec un écart type d'environ 93 mg/dl. Cette moyenne se situe dans la zone désirable du taux de cholestérol qui doit être inférieur à 200 mg/dl. De plus, un individu sur quatre a au plus 173,25 mg/dl tandis qu'un individu sur deux a un taux médian de 223 mg/dl ce qui est relativement au-dessus de la norme avec plus de 23 mg/dl.

4. Analyse et interprétation du rythme cardiaque maximum

La fréquence cardiaque maximale (FCM) est le rythme que le cœur d'une personne donnée atteint lors des plus fortes sollicitations. Au repos, la fréquence cardiaque est environ de 60 à 80 battements



par minute.

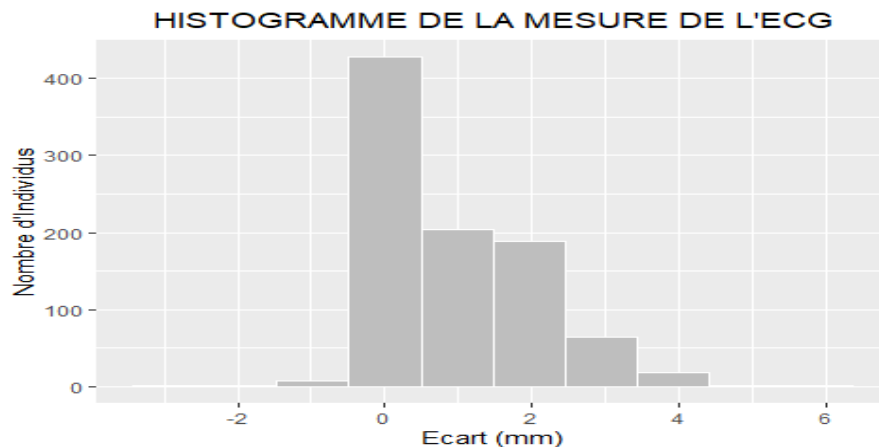
Le rythme cardiaque des individus soumis à notre étude est **compris entre 60 et 202 battements par minute**, alors que la valeur modale est de 150. Ce qui signifie que **la grande majorité de ces personnes ont un rythme cardiaque qui s'élève à 150 battements par minute**.

Le nombre de battement moyen par minute du cœur est de 137, avec un écart type d'environ 25 battements par minute.

De plus, un individu sur quatre a au plus 120 battements par minute, tandis qu'une personne sur deux a 138 battements par minute.

5. Analyse et interprétation de la mesure sur l'électrocardiogramme

La mesure sur l'électrocardiogramme (ECG) appelée "Oldpeak" est généralement exprimée en millimètres et représente *la différence entre le point le plus bas du segment ST pendant l'exercice et la ligne de base (point de départ) de l'ECG*. C'est un indicateur de la réponse électrique du cœur pendant l'effort et peut être utilisé pour évaluer la présence d'ischémie cardiaque, ce qui peut être un



signe de maladie coronarienne. **Une valeur élevée peut être un indicateur de risque de maladie cardiaque.**

La mesure sur l'électrocardiogramme oscille entre -2.6 mm et 6.2 mm, alors que la valeur modale est d'environ 0.02. Ce qui signifie que la plupart de ces personnes ont un Oldpeak qui s'élève à 0.02 mm, cela indique que **la majorité des individus n'ont pas un risque de maladie cardiaque élevé car cette valeur est très faible.**

La moyenne des individus a une valeur égale à 0.88 mm, avec un écart type d'environ 1.06 mm. C'est à dire que la mesure du segment est comprise entre -0,12 mm et 2,54 mm.

De plus, un individu sur quatre a au plus 0 mm, ce qui montre que **le quart des individus ont un bon niveau**, tandis qu'**une personne sur deux a 1 mm**, ce qui n'est pas très grave.

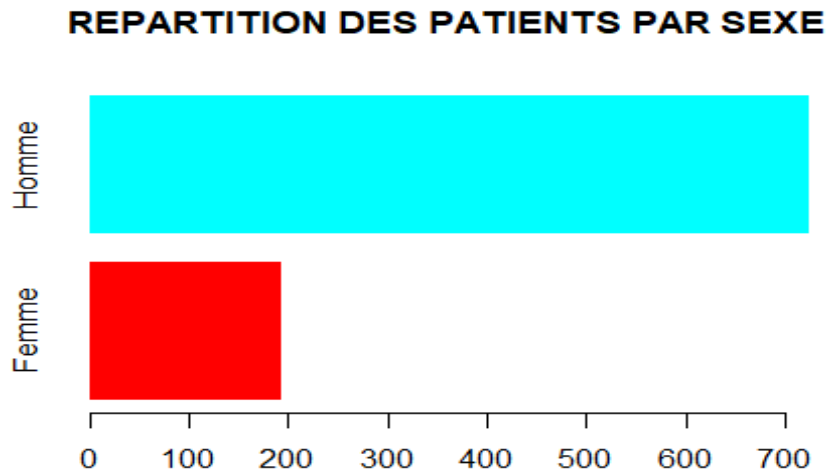
L'histogramme et la valeur du coefficient d'asymétrie (1) montrent que la distribution est étalée à droite.

II. ANALYSE DES VARIABLES QUALITATIVES

Dans cette partie, il s'agira de faire des graphiques, de calculer les indicateurs adaptés et de les interpréter de toutes les variables qualitatives de notre jeu de données.

1. Analyse de la variable Sexe

Le diagramme ci-dessous permet de mieux visualiser la répartition des patients par sexe.



Il nous permet de constater que notre jeu de données est composé majoritairement d'homme.

En se basant sur les résultats du tableau ci-dessous nous sommes en mesure d'affirmer que 79% d'individus sont de sexe masculin et que 21% sont de sexe féminin.

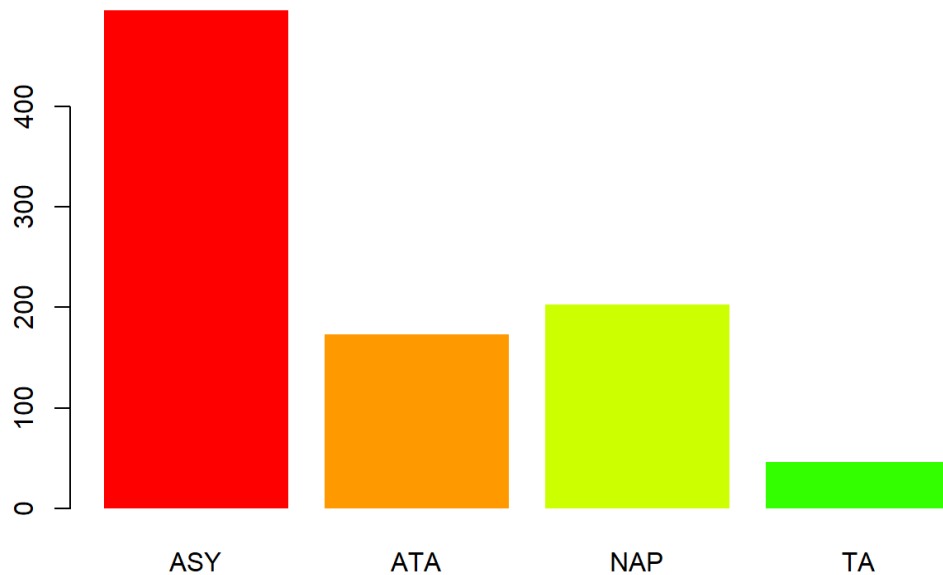
Tableau 6: Tableau Statistique de l'âge des patients

	Effectif	Fréquence
Femme	193	0.2102397
Homme	725	0.7897603

2. Analyse de la variable *ChestPaintype*

La taille des barres met en exergue que la majorité des patients ont une douleur thoracique de type asymptomatique (ASY), c'est à dire qu'ils ne manifestent pas de symptôme. Ensuite, viennent les patients qui ont d'autres type de douleur (NAP) suivi de ceux qui ont des douleurs de type atypique (ATA), en d'autres termes, ces douleurs sont différentes du type habituel. Enfin, c'est seulement une

REPARTITION DES PATIENTS PAR TYPE DE DOULEUR THORACIQUE



minorité qui présente des douleurs de poitrine typique.

Le tableau statistique ci-dessous nous donne les proportions suivantes :

- 54% des patients ne présentent pas de symptôme d'angine de poitrine.
- Concernant l'Angine de poitrine atypique (ATA), 18,84% des individus en souffre.
- Enfin, les patients présentant l'Angine de poitrine (TA) et des Douleurs autres que l'angine (NAP) sont respectivement de 5% et 22,11%.

Tableau 7: Tableau Statistique du type d'angine de poitrine

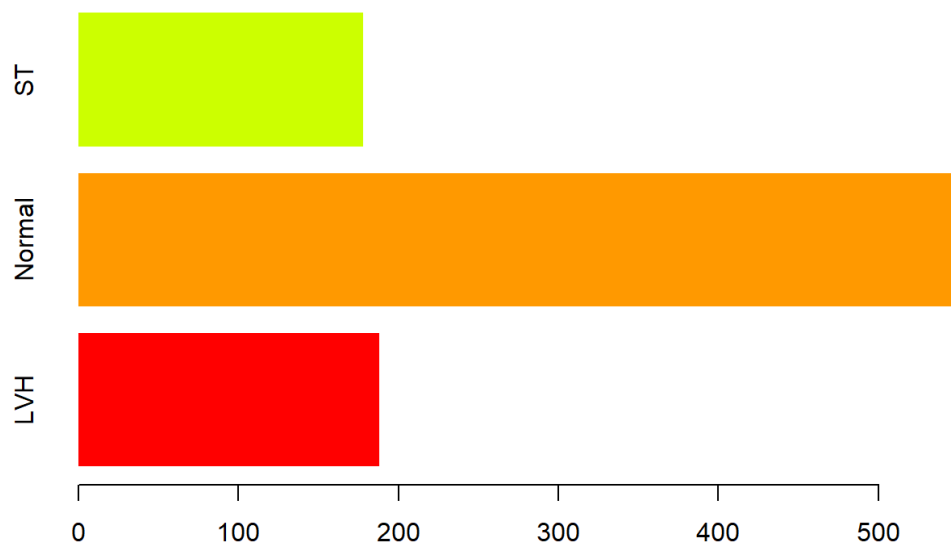
	Effectif	Frequence
ASY	496	0.5403050
ATA	173	0.1884532
NAP	203	0.2211329
TA	46	0.0501089

En résumé, la majorité des patients ne présentent pas de symptôme d'angine de poitrine.

3. Analyse de la variable *RestingECG*

Le graphique nous permet de constater que la plupart des patients sont dans une situation normale suite au test de l'ECG au repos.

REPARTITION DES PATIENTS PAR RESULTAT DE ECG



Le tableau statistique ci-dessous nous confirme la conclusion de l'observation du graphique avec plus de 60% des patients qui ont un rythme cardiaque normal.

Tableau 8: Tableau Statistique ECG au repos

	Effectif	Fréquence
LVH	188	0.2047930
Normal	552	0.6013072
ST	178	0.1938998

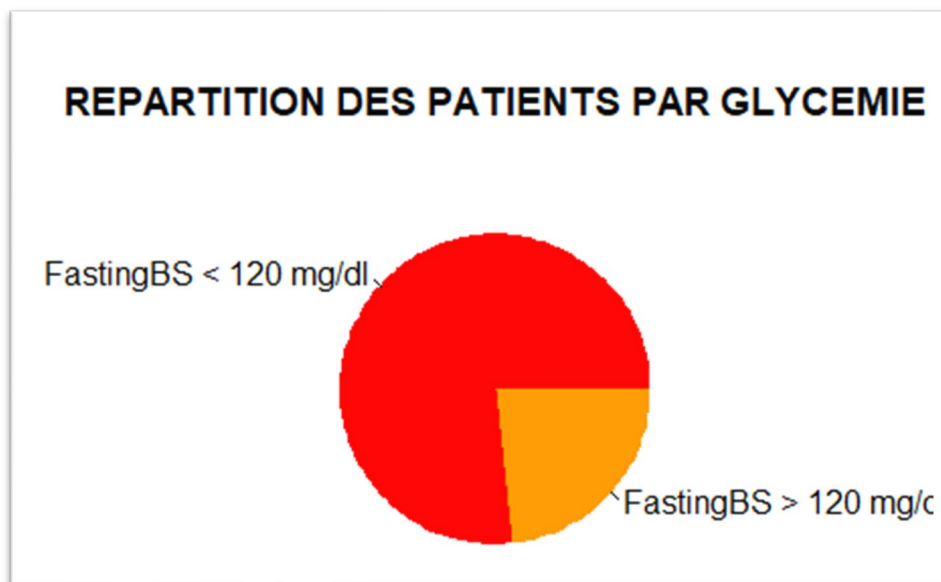
De plus, 20,47 % présentent une hypertrophie ventriculaire gauche (LVH) qui peut être associée à des problèmes de santé tels que des maladies cardiaques, des valvulopathies, ou d'autres conditions médicales.

Enfin, 19,39% ont été diagnostiqués d'une anomalie au niveau du segment ST-T ce qui pourrait avoir des raisons diverses :

- Ischémie myocardique qui est une diminution de l'apport sanguin au muscle cardiaque
- Infarctus du myocarde ou crise cardiaque
- Angine de poitrine stable ou instable dû à une dépression du segment ST ou une élévation

4. Analyse et interprétation de la glycémie à jeun.

Le taux de glycémie à jeun est la concentration de glucose (sucre) dans le sang après une période de jeûne, généralement d'au moins 8 heures. Cette mesure est souvent utilisée pour évaluer le métabolisme du glucose et le risque de diabète.



Il est aisément visible sur la représentation graphique que plus des deux-tiers (2/3) des patients ont une glycémie à jeun inférieure à 120 mg/dl.

Le tableau des fréquences ci-dessous permet de quantifier cette observation.

Le taux d'individus ayant leur *glycémie à jeun supérieur à 120 mg/dl* est de 23%, c'est à dire que ces derniers ont le *Diabète*. Plus de 77% ont un *taux inférieur à 120 mg/dl*, ce qui signifie qu'ils sont soit dans la zone de *Prédiabète* ou dans la zone *Normale*.

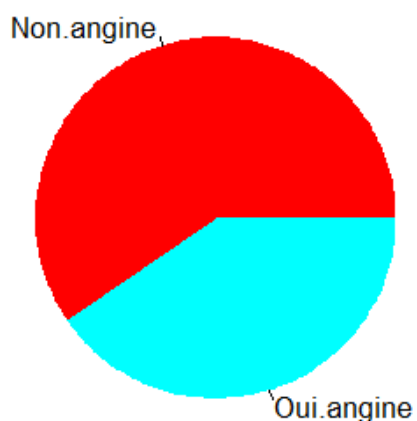
Tableau 9:Tableau Statistique de la glycémie à jeun des patients

	Effectif	Frequence
FastingBS < 120 mg/dl	704	0.7668845
FastingBS > 120 mg/dl	214	0.2331155

5. Analyse et interprétation de l'Angine induite par L'exercice.

L'angine induite par l'exercice, également appelée angine d'effort, est une douleur thoracique ou un inconfort qui se produit pendant l'activité physique ou peu de temps après. Elle est généralement causée par un apport sanguin insuffisant au muscle cardiaque (ischémie) en raison d'une maladie coronarienne sous-jacente.

REPARTITION DES ANGINES INDUITES PAR EXERCICE



Le diagramme circulaire et le tableau statistique nous permettent d'affirmer sans doute que parmi nos individus, 59,59 % *n'ont pas une angine induite par l'exercice*, c'est à dire qu'ils ne ressentent pas de douleur thoracique ou d'inconfort qui se produit pendant l'activité physique ou peu de temps après.

Cependant, **40,41 % reconnaissent avoir une angine induite par l'exercice.**

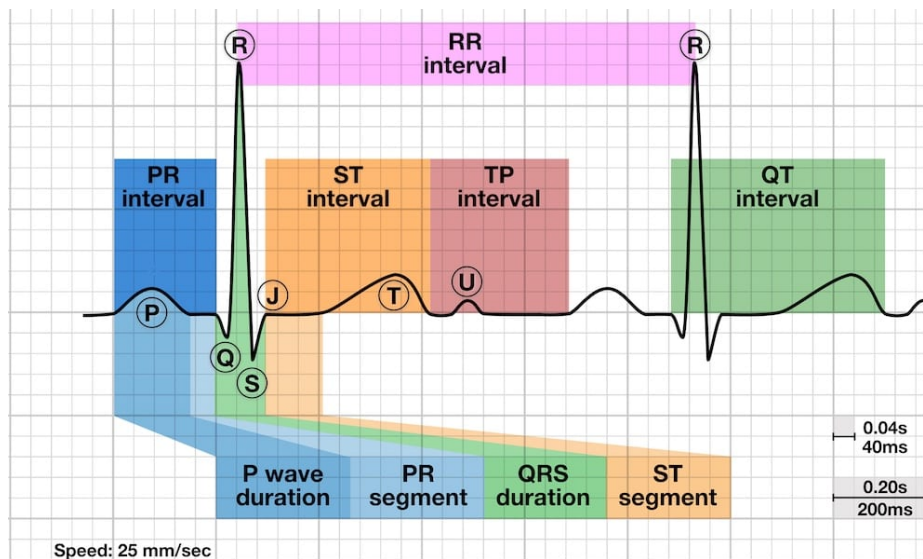
Tableau 10: Tableau Statistique de l'angine après Exercice

	Effectif	Fréquence
Non.angine (Présence)	547	0.5958606
Oui.angine (Absence)	371	0.4041394

6. Analyse et interprétation de la forme du Segment ST.

L'interprétation de la pente du segment ST sur un électrocardiogramme (ECG) est importante pour évaluer la présence d'anomalies cardiaques.

Figure 3: Tracé de l'ECG avec les Intervalles



Le segment ST est la portion du tracé de l'ECG située entre la fin du complexe QRS et le début de l'onde T.

La pente du segment ST peut être classée en trois catégories principales : ascendant (positif), descendant (négatif), ou horizontal (isoélectrique).

Elles sont représentées dans notre jeu de donnée, respectivement par : Up, Down et Flat.

Tableau 11: Tableau statistique de la forme du Segment ST

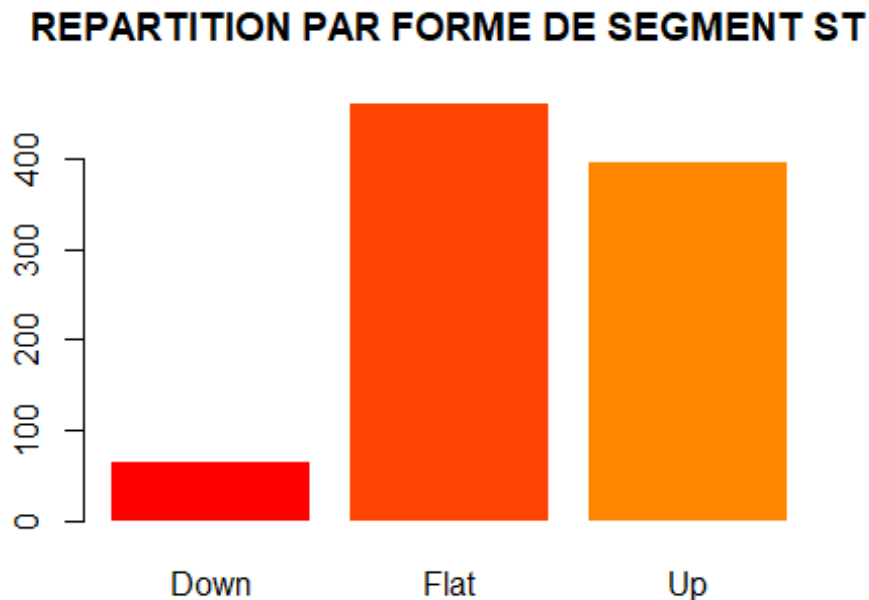
	Effectif	Fréquence
Down	63	0.0686275
Flat	460	0.5010893
Up	395	0.4302832

Parmi nos individus, plus de 50 % ont une forme de segment ST plate, ce qui indique une situation normale de la santé cardiaque.

Cependant, 43 % ont une courbe ascendante, indiquant une ischémie myocardique, surtout si elle est accompagnée de symptômes tels que la douleur thoracique. Cette forme de la courbe peut être observée dans des conditions d'angine de poitrine ou pendant un test d'effort.

Seulement 7 % des courbes ont une forme descendante (Down), cela peut être observé lors d'un infarctus du myocarde (crise cardiaque).

Cette conclusion est aisément visible sur le graphique ci-dessous :



7. Analyse et interprétation de la présence ou pas de maladie cardiaque.

Les maladies cardiaques, également connues sous le nom de maladies cardiovasculaires, englobent un large éventail de troubles affectant le cœur et les vaisseaux sanguins.

Ces affections peuvent avoir des causes diverses, notamment des facteurs génétiques, des modes de vie malsains et d'autres conditions médicales.

Le diagramme ci-dessous nous permet de visualiser la présence ou l'absence de maladie cardiaque chez les patients.

Nous pouvant constater qu'un peu plus de la moitié des patients ont une maladie cardiaque, les chiffres du tableau nous permettront de savoir la répartition chiffrée des patients en fonction de la présence ou non de maladie cardiaque.

REPARTITION PAR PRESENCE DE MALADIE



Tableau 12: Tableau statistique des maladies cardiaques

	Effectif	Fréquence
Non	410	0.4466231
Oui	508	0.5533769

De façon chiffrée, **plus de 55 % des patients souffre de maladie du coeur tandis que 44 % n'en souffre pas.**

Conclusion

Au terme de cette partie, nous retenons que notre jeu de donnée est composé de **918 patients** qui sont à **79% des hommes adultes.**

Ces patients sont à **23% des malades de diabète** et à **55% des malades cardiaques.**

Dans la suite de notre analyse, il s'agira de déterminer les facteurs qui sont liés aux maladies cardiaques et de montrer l'intensité de ces liaisons.

TROISIEME PARTIE : ANALYSE BIVARIEE

Introduction

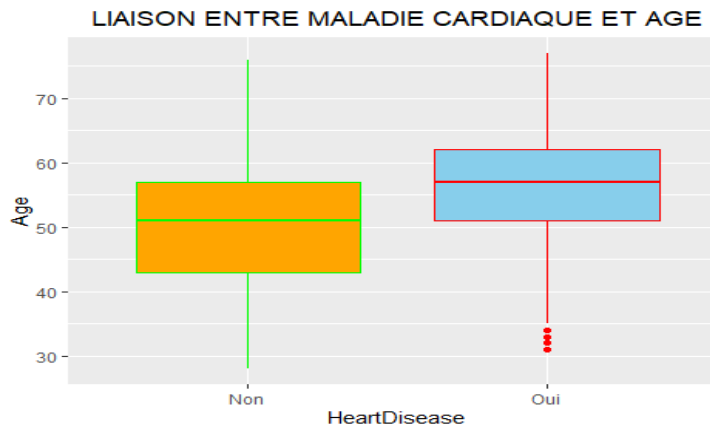
Dans Cette dernière partie, nous allons évaluer l'effet des onze autres variables sur les maladies du coeur (HeartDisease).

Par ailleurs, nous déterminerons à l'aide des indicateurs, les facteurs qui sont les plus liés aux maladies cardiaques grâce aux indicateurs de corrélation.

I. Analyse bivariée des variables quantitatives

1. Les maladies cardiaques sont-elles influencées par l'âge ?

a. Analyse graphique



La représentation ci-dessus nous montre l'existence d'une liaison entre la présence ou non de maladies du coeur et l'âge des patients.

Afin d'avoir une idée plus claire de cette situation, nous allons calculer le rapport de corrélation.

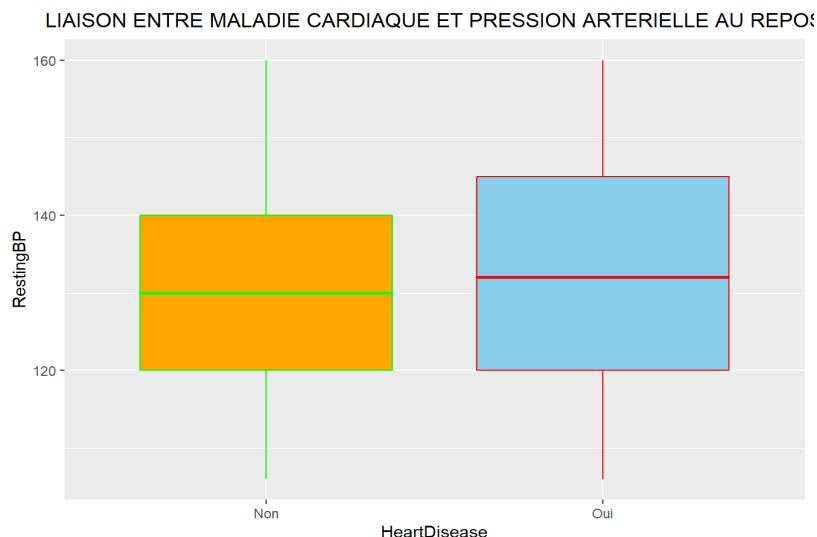
b. Analyse par des indicateurs

\$Rapport corrélation						
[1] 0.07954572						
\$Resultat_Test_Anova						
Response: vecteur						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
facteur	1	6490	6490.1	79.161	< 2.2e-16 ***	
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1						
\$Anova.P.value						
[1] 3.007953e-18						
\$Significativite_TestAnova						
[1] "liaison significative, les deux variables sont liées"						
\$Intensite_liaison						
[1] "liaison très faible"						

Suite au calcul de ce rapport, **il ressort qu'il existe une liaison entre ces deux variables**, cependant, la valeur du rapport de corrélation (**0.079**) indique une *intensité de liaison très faible*. Par conséquent, **les maladies cardiaques sont expliquées dans de faibles proportions par l'âge des personnes.**

2. Les maladies cardiaques sont-elles influencées par la Pression artérielle au repos ?

a. Analyse graphique



La représentation ci-dessus nous autorise à soupçonner l'existence d'une liaison entre la présence des maladies du coeur et les valeurs de la pression cardiaque au repos des patients.

Cela s'explique par le fait que nous constatons une variation de tailles de la boîte en fonction de l'existence ou non de maladie cardiaque. La médiane au niveau de ceux qui ont la maladie est un peu plus au-dessus de ceux qui ne l'ont pas.

Ces variations montrent la probable relation entre les deux faits. Afin d'avoir une idée plus claire de cette situation, nous allons calculer le rapport de corrélation.

b. Analyse par des indicateurs

\$Rapport_Correlation
[1] 0.01338686
\$Resultat_Test_Anova
Analysis of Variance Table
Response: vecteur
\$Anova.P.value
[1] 0.0004437227
\$Significativite_TestAnova
[1] "liaison significative, les deux variables sont liées"
\$Intensite_liaison
[1] "liaison très faible"

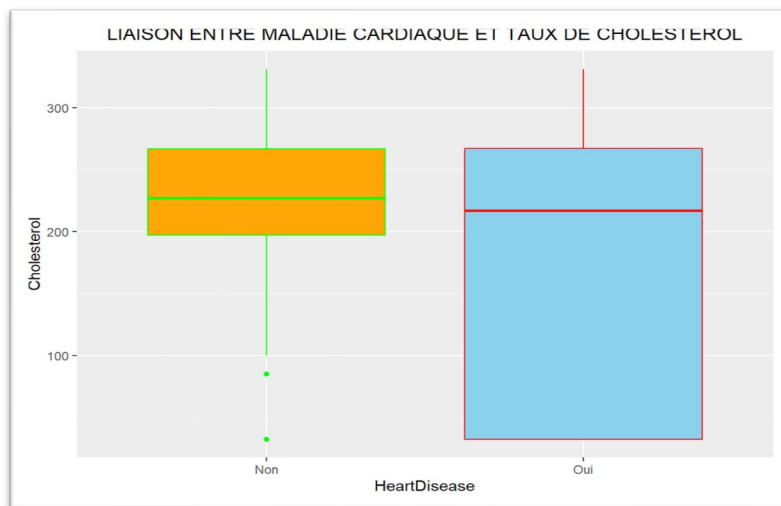
Suite au calcul de ce rapport, il ressort qu'il existe une liaison entre ces deux variables, cependant, *l'intensité de cette liaison est très faible* avec une valeur du rapport de corrélation proche de (0).

Par conséquent, ***le niveau de la pression artérielle au repos explique très faiblement l'existence d'une maladie cardiaque.***

3. Les maladies cardiaques sont-elles influencées par le taux de Cholestérol ?

a. Analyse graphique

La représentation ci-dessus nous autorise à soupçonner l'existence d'une liaison entre la présence des maladies cardiaques et le taux de Cholestérol des patients. Nous constatons une variation de tailles de la boîte en fonction de l'existence ou non de la maladie.



La médiane au niveau de ceux qui ont la maladie est légèrement en dessous de ceux qui n'en n'ont pas et le nombre d'individus dont le taux est inférieur à la médiane augmente drastiquement (voir la boîte à moustache en Bleu).

b. Analyse par des indicateurs

Afin d'avoir une idée plus claire de cette situation, nous allons calculer le rapport de corrélation.

\$Rapport_Correlation

[1] 0.05490435

\$Resultat_Test_Anova

Analysis of Variance Table

\$Anova.P.value

[1] 6.474056e-13

\$Significativite_TestAnova

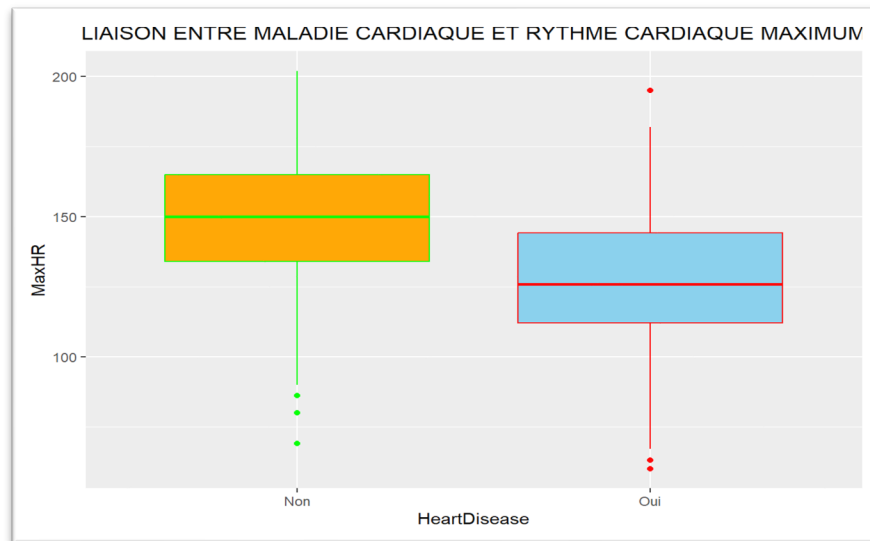
[1] "liaison significative, les deux variables sont liées"

\$Intensite_liaison

[1] "liaison très faible"

Suite au calcul de cet indicateur donne une valeur égale à 0.055. Il ressort qu'il existe une liaison entre ces deux variables, avec *une intensité très faible*. En claire, ***le niveau du taux de cholestérol explique faiblement l'existence d'une maladie cardiaque.***

4. Les maladies cardiaques sont-elles influencées par le rythme cardiaque maximum ?



a. Analyse graphique

La représentation ci-dessus nous permet de constater l'existence d'une liaison entre la présence de maladies cardiaques et le niveau maximum du rythme cardiaque des patients.

Nous constatons un changement de position de la boîte en fonction de l'existence ou non de maladie cardiaque. La médiane au niveau de ceux qui ont la maladie est un peu plus basse que celle de ceux ne l'ont pas.

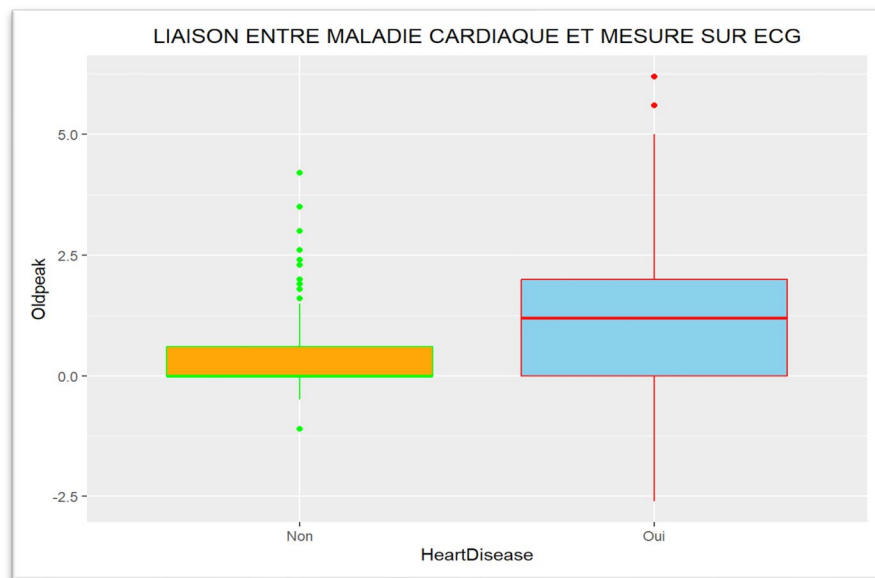
b. Analyse par des indicateurs

Afin d'avoir une idée plus claire de cette situation, nous allons calculer le rapport de corrélation.

\$Rapport_Correlation
[1] 0.1603368
\$Resultat_Test_Anova
Analysis of Variance Table
\$Anova.P.value
[1] 1.137786e-36
\$Significativite_TestAnova
[1] "liaison significative, les deux variables sont liées"
\$Intensite_liaison
[1] "Liaison faible"

Suite au calcul de ce rapport, il ressort qu'il existe une liaison entre ces deux variables, cependant, *l'intensité de cette liaison est faible* avec une valeur du rapport de corrélation égale à 0.16. Par conséquent, *le rythme cardiaque maximum explique faiblement l'existence d'une maladie cardiaque.*

5. Les maladies cardiaques sont-elles influencées par la mesure sur l'électrocardiogramme (ECG) ?



a. Analyse graphique

La représentation ci-dessus nous autorise à soupçonner l'existence d'une liaison entre la présence des maladies cardiaques et le niveau maximum du rythme cardiaque des patients. Nous constatons un changement de position et de taille des boîtes en fonction de l'existence ou non de la maladie.

La médiane au niveau de ceux qui ont la maladie est un peu plus basse que celle de ceux ne l'ont pas.

b. Analyse par des indicateurs

Afin d'avoir une idée plus claire de cette situation, nous allons calculer le rapport de corrélation.

\$Rapport_correlation

[1] 0.1631762

\$Resultat_Test_Anova

\$Anova.P.value

[1] 2.390772e-37

\$Significativite_TestAnova

[1] "liaison significative, les deux variables sont liées"

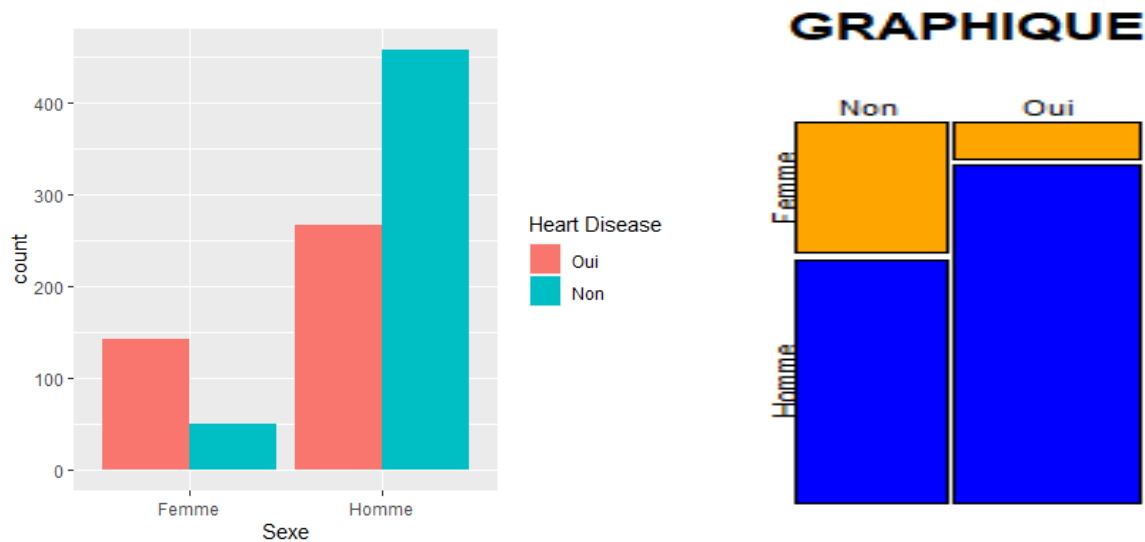
\$Intensite_liaison

[1] "liaison faible"

Suite au calcul de ce rapport, il ressort qu'il existe une liaison entre ces deux variables, cependant, *l'intensité de cette liaison est faible* avec une valeur du rapport de corrélation égale à 0.16. Par conséquent, *la mesure sur l'électrocardiogramme explique faiblement l'existence d'une maladie cardiaque.*

II. Liaison entre Les maladies cardiaques et les variables qualitatives

1. Les maladies cardiaques sont-elles influencées par le genre ?



a. Analyse graphique

Débutons par visualiser la répartition des maladies du coeur par sexe grâce à des diagrammes.

Ces graphiques nous donnent une idée de la liaison entre ces deux variables : *les individus de sexe masculin semblent être plus sensible aux maladies cardiaques.* Dans le but de clarifier cette probable relation, nous allons analyser le tableau des fréquences, le tableau des profils et calculer les indicateurs de liaison qui sont le V de Cramer et le Khi-deux.

b. Analyse par des indicateurs

TABLEAU DE CONTINGENCE

	Femme	Homme
Non	143	267
Oui	50	458

TABLEAU DE FREQUENCE

	Femme	Homme
Non	0.35	0.65
Oui	0.10	0.90

	Femme	Homme
Non	0.74	0.37
Oui	0.26	0.63

	Femme	Homme
Non	0.16	0.29
Oui	0.05	0.50

L'analyse de ces tableaux montre que **50% de l'ensemble des individus sont de sexe masculin et ont une maladie cardiaque** tandis 5% seulement sont des femmes et sont malades.

\$Khi_Deux
X-squared
84.1451
\$V_Cramer
[1] 0.3054449
\$Khi2.P.value
[1] 4.597617e-20
\$Significativite_TestKhi2
[1] "liaison significative, les deux variables sont liees"
\$Fisher.P.value
[1] 1.769293e-20
\$Significativite_TestFisher
[1] "liaison significative, les deux variables sont liees"
\$Intensite_liaison
[1] "liaison faible"

De plus, le tableau des profils lignes montre que **90% des malades sont des hommes, tandis que 10% seulement sont des femmes**. Ce qui renforce nos résultats soupçonnés au niveau de l'analyse graphique.

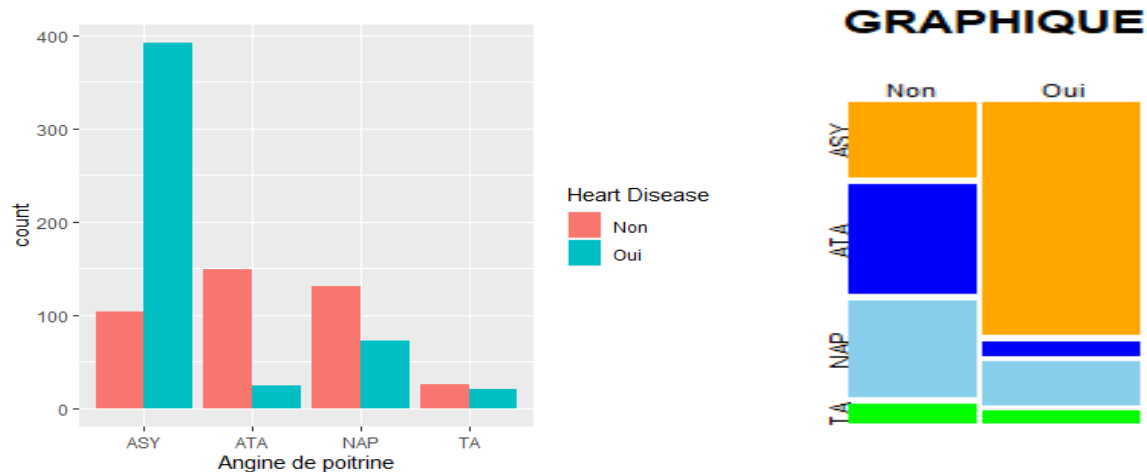
Le calcul des indicateurs de corrélation ci-dessus viennent confirmer et renforcer la pertinence des résultats obtenus précédemment.

Par conséquent, *il existe une liaison significative mais faible* entre l'existence des maladies cardiaques et le genre des individus.

2. Les maladies cardiaques sont-elles influencées par l'Angine de poitrine ?

a. Analyse graphique

Débutons par visualiser la répartition des maladies du cœur par Angine de poitrine.



Ces graphiques nous donnent une idée de la liaison entre ces deux variables :

Les individus ne présentant pas de symptôme d'angine de poitrine semblent être plus sensible aux maladies cardiaques. Dans le but de clarifier cette probable relation, nous allons analyser les tableaux de contingence, de fréquence et des profils et calculer les indicateurs de liaison qui sont le V de Cramer et le Khi-deux.

b. Analyse par des indicateurs

* Tableau statistique

TABLEAU DE CONTINGENCE

	ASY	ATA	NAP	TA
Non	104	149	131	26
Oui	392	24	72	20

TABLEAU DE FREQUENCE

	ASY	ATA	NAP	TA
Non	0.11	0.16	0.14	0.03
Oui	0.43	0.03	0.08	0.02

TABLEAU DES PROFILS LIGNES

	ASY	ATA	NAP	TA
Non	0.25	0.36	0.32	0.06
Oui	0.77	0.05	0.14	0.04

TABLEAU DES PROFILS COLONNES

	ASY	ATA	NAP	TA
Non	0.21	0.86	0.65	0.57
Oui	0.79	0.14	0.35	0.43

L'analyse de ces deux tableaux montre que *43% de l'ensemble des patients ont une maladie du coeur sans symptôme d'angine de poitrine.*

Aussi, 8% de l'ensemble des patients ont une maladie du coeur accompagnée de types de douleurs autres que celles dû à l'angine de poitrine (NAP). En ce qui concerne la relation entre les maladies cardiaques et les douleurs dû à l'angine de poitrine, seulement 3% des patients présentent cette maladie accompagnée d'une angine de poitrine atypique (ATA) et 2% d'une angine de poitrine typique (AT).

Cela représente respectivement 5% et 4% des individus malades cardiaques. Alors que ceux n'ayant *pas de symptôme représentent 77% des malades* et ceux ayant des douleurs autre que l'angine ont une proportion de 14% des malades cardiaques.

* *Indicateur de liaison*

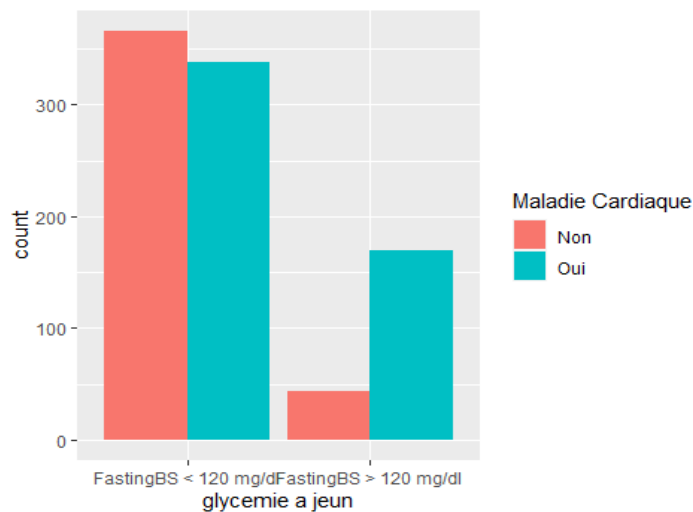
\$Khi_Deux
X-squared
268.0672
\$V_Cramer
[1] 0.5403816
\$Khi2.P.value
[1] 8.083728e-58
\$Significativite_TestKhi2
[1] "liaison significative, les deux variables sont liées"
\$Fisher.P.value
[1] 4.164861e-62
\$Significativite_TestFisher
\$Intensite_liaison
[1] "liaison moyenne"

Le calcul des indicateurs de corrélation de Khi-deux et le V de Cramer permettent de déterminer la vraie nature de la liaison entre les maladies du cœur et l'Angine de poitrine.

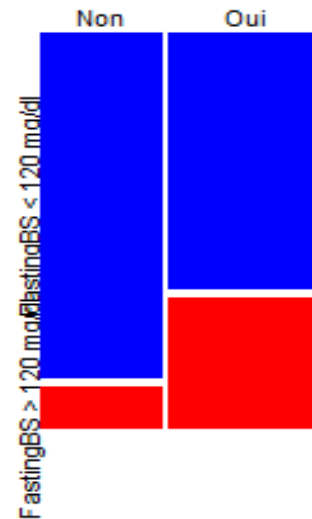
Par conséquent, *il existe une liaison significative d'intensité moyenne* entre l'existence des maladies cardiaques et l'Angine de poitrine.

3. Les maladies cardiaques sont-elles influencées par la glycémie à jeun ?

a. Analyse graphique



GRAPHIQUE



Débutons par visualiser la répartition des maladies du coeur par rapport à la glycémie à jeun.

Ces graphiques nous donnent un aperçu de la liaison entre ces deux variables. Tout d'abord, FastingBS > 120 mg/dl signifie que la glycémie à jeun est supérieur a 120 mg/dl ce qui est synonyme de présence de diabète chez le patient.

Certes, la visualisation des graphiques ci-dessus montre que la majorité des personnes ne présentant pas de maladie cardiaque n'ont pas de diabète (glycémie à jeun est inferieur a 120 mg/dl), mais la proportion des malades cardiaques ayant diabète est plus grande que celle des personnes saines ayant le diabète, ce qui nous permet de soupçonner que : *les individus ayant le diabète semblent être plus sensible aux maladies cardiaques.*

Dans le but de clarifier cette probable relation, nous allons analyser le tableau de contingence, le tableau de fréquence et des profils et calculer les indicateurs de liaison qui sont le V de Cramer et le Khi-deux.

b. Analyse par des indicateurs

* Tableau statistique

TABLEAU DE CONTINGENCE

	FastingBS <120 mg/dl	FastingBS>120 mg/dl
Non	366	44
Oui	338	170

TABLEAU DE FREQUENCE

	FastingBS < 120 mg/dl	FastingBS > 120 mg/dl
Non	0.40	0.05
Oui	0.37	0.19

TABLEAU DES PROFILS LIGNES

	FastingBS < 120 mg/dl	FastingBS > 120 mg/dl
Non	0.89	0.11
Oui	0.67	0.33

TABLEAU DES PROFILS COLONNES

	FastingBS < 120 mg/dl	FastingBS > 120 mg/dl
Non	0.52	0.21
Oui	0.48	0.79

L'analyse de ces tableaux montre que 37% de l'ensemble des patients ont une maladie du coeur et n'ont pas le diabète. Aussi, seulement 5% de l'ensemble des patients ont le diabète sans avoir de maladie cardiaque alors que 19% des individus présentent les deux maladies.

L'analyse du tableau des profils colonnes met en lumière que *79 diabétiques sur cent ont une maladie cardiaque*, ce qui est très évocateur et nécessite un approfondissement afin d'en cerner les tenants et aboutissants.

- Indicateur de liaison

\$Khi_Deux
X-squared
64.32068
\$V_Cramer
[1] 0.2672912
\$Khi2.P.value
[1] 1.057302e-15
\$Significativite_TestKhi2
[1] "liaison significative, les deux variables sont liees"
\$Fisher.P.value
[1] 1.041182e-16
\$Significativite_TestFisher
\$Intensite_liaison
[1] "liaison faible"

Le calcul des indicateurs de corrélation de Khi-deux et le V de cramer permettent déterminer la vraie nature de la liaison entre les maladies du coeur et le diabète.

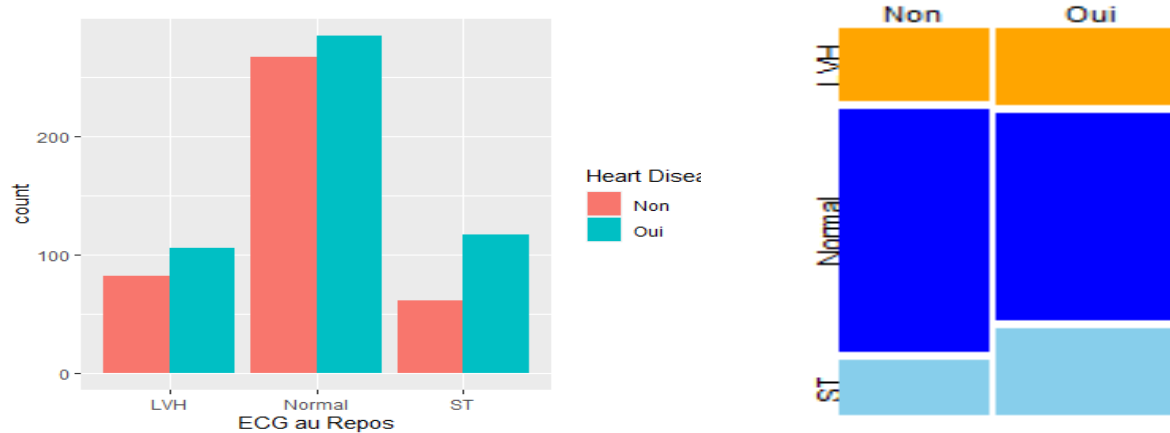
Par conséquent, *il existe une liaison significative d'intensité faible* entre l'existence des maladies cardiaques et le diabète.

Mais, la valeur du V de cramer (0.2672912) est relativement élevé par rapport a celle des autres variables calculées précédemment.

4. Les maladies cardiaques sont-elles influencées par le résultat de l'électrocardiogramme ?

a. Analyse graphique

Débutons par visualiser la répartition des maladies du coeur par rapport au résultat de l'électrocardiogramme grâce à des diagrammes en bar.



Ces graphiques nous montrent que la grande majorité des malades présentent une activité électrique du cœur se trouvant dans les limites normales, sans anomalies significatives (Barres du milieu).

Cependant, les barres de droite et de gauche sont synonymes de maladies liées au coeur. Les patients présentant une anomalie au niveau du segment ST et ayant une maladie sont relativement élevés par rapport à ceux qui n'ont pas la maladie et qui ont une anomalie, ce raisonnement est applicable aussi au LVH.

Dans le but de clarifier cette probable relation, nous allons analyser le tableau de contingence, le tableau de fréquence et des profils et calculer les indicateurs de liaison qui sont le V de Cramer et le Khi-deux.

b. Analyse par des indicateurs

** Tableau statistique*

TABLEAU DE CONTINGENCE

	LVH	Normal	ST
Non	82	267	61
Oui	106	285	117

	LVH	Normal	ST
Non	0.09	0.29	0.07
Oui	0.12	0.31	0.13

TABLEAU DES PROFILS LIGNES

	LVH	Normal	ST
Non	0.20	0.65	0.15
Oui	0.21	0.56	0.23

TABLEAU DES PROFILS COLONNES

	LVH	Normal	ST
Non	0.44	0.48	0.34
Oui	0.56	0.52	0.66

Le tableau des fréquences montre que 31% des patients ont une maladie cardiaque et se situent dans la zone normale de la mesure de l'ECG.

Relativement aux patients ayant une Hypertrophie du ventricule gauche (LVH) et une anomalie du segment ST accompagnés de maladie cardiaque, leur proportion est respectivement de 12% et 13%.

Par ailleurs, 56% de ceux qui ont une Hypertrophie du ventricule gauche (LVH) ont une maladie cardiaque. Egalement, 66% des patients ayant une anomalie au niveau du segment ST sont des malades cardiaques. Aussi, 52% des patients qui ont un ECG normal sont malades.

** Indicateur de liaison*

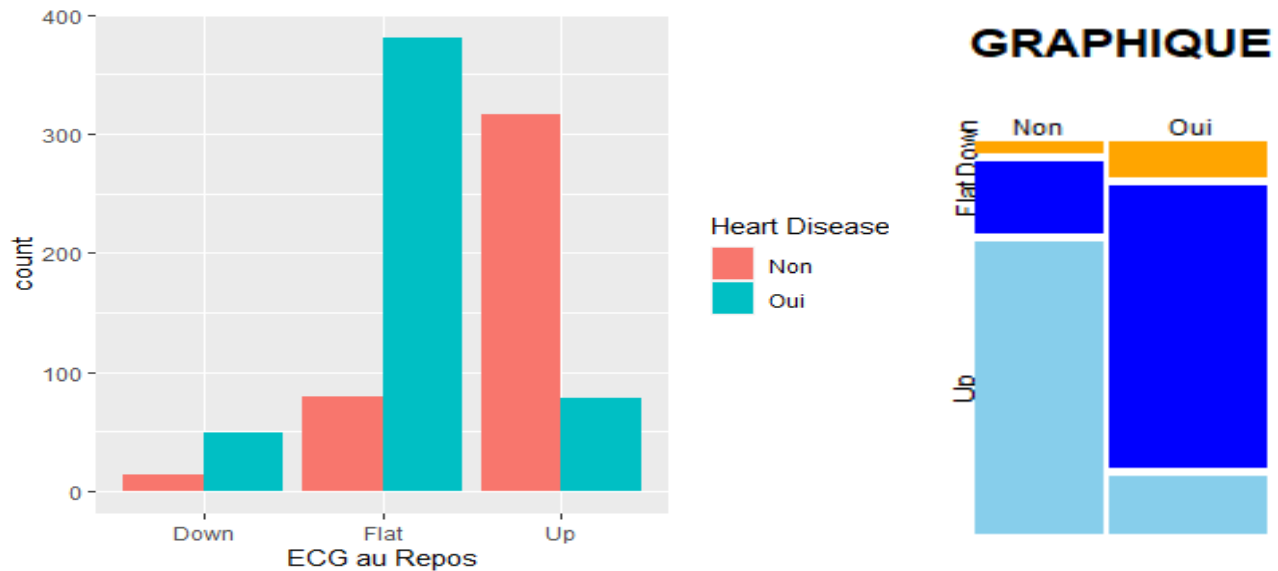
\$Khi_Deux
X-squared
64.32068
\$V_Cramer
[1] 0.2672912
\$Khi2.P.value
[1] 1.057302e-15
\$Significativite_TestKhi2
[1] "liaison significative, les deux variables sont liées"
\$Fisher.P.value
[1] 1.041182e-16
\$Significativite_TestFisher
\$Intensite_liaison
[1] "liaison faible"

Le calcul des indicateurs de corrélation de Khi-deux et le V de cramer permettent déterminer la vraie nature de la liaison entre les maladies du coeur et l'Angine de poitrine. Par conséquent, *il existe une liaison significative d'intensité faible* entre l'existence des maladies cardiaques et la nature de la mesure sur l'ECG.

5. Les maladies cardiaques sont-elles influencées par la pente de l'ECG ?

a. Analyse graphique

Débutons par visualiser la répartition des maladies du coeur par rapport au résultat de l'électrocardiogramme grâce à un diagramme en bar.



Ces graphiques nous montrent que **la grande majorité des malades ont une pente du segment ST de forme plate (voir les barres du milieu)**. Ce qui dénote que les malades qui ont une courbe plate sur l'ECG ont des maladies cardiaques.

Cependant, les barres de droite et de gauche sont synonymes de maladies liées au coeur. Les patients présentant *Une élévation du segment ST* peuvent être associées à des conditions telles qu'un infarctus du myocarde.

Aussi, les individus qui ont une *dépression du segment ST* peuvent être associées à diverses conditions, y compris une ischémie myocardique (manque d'apport sanguin au muscle cardiaque).

Dans le but de clarifier cette probable relation, nous allons analyser le tableau de contingence, le tableau de fréquence et des profils et calculer les indicateurs de liaison qui sont le V de Cramer et le Khi-deux.

b. Analyse par des indicateurs

** Tableau statistique*

Tableau De Contingence

	Down	Flat	Up
Non	14	79	317
Oui	49	381	78

Tableau De Fréquence

	Down	Flat	Up
Non	0.02	0.09	0.35
Oui	0.05	0.42	0.08

Tableau Des Profils Colonnes

	Down	Flat	Up
Non	0.22	0.17	0.8
Oui	0.78	0.83	0.2

L'analyse des tableaux ci-dessus montre que **42% de l'ensemble des patients ont une maladie du coeur et une forme du segment ST normale (plate).**

De plus, 8% de l'ensemble des patients ont une maladie du coeur avec une forme ascendante du segment ST. Nous pouvons aisément remarquer dans le tableau des profils colonnes, que 83% des individus qui ont une pente plate du segment ST ont des maladies cardiaques.

Par ailleurs, **78% de ceux qui ont une dépression du segment ST ont une maladie cardiaque** et seulement **20% des patients qui ont une élévation du segment ST ont une maladie cardiaque.**

Ces derniers résultats sont associés a des maladies telles que : l'infarctus du myocarde ou une ischémie myocardique (manque d'apport sanguin au muscle cardiaque).

** Indicateur de liaison*

\$Khi_Deux
X-squared
355.9184
\$V_Cramer
[1] 0.6226642
\$Khi2.P.value
[1] 5.167638e-78
\$Significativite_TestKhi2
[1] "liaison significative, les deux variables sont liées"
\$Fisher.P.value
[1] 9.583533e-84
\$Significativite_TestFisher
\$Intensite_liaison
[1] "liaison forte"

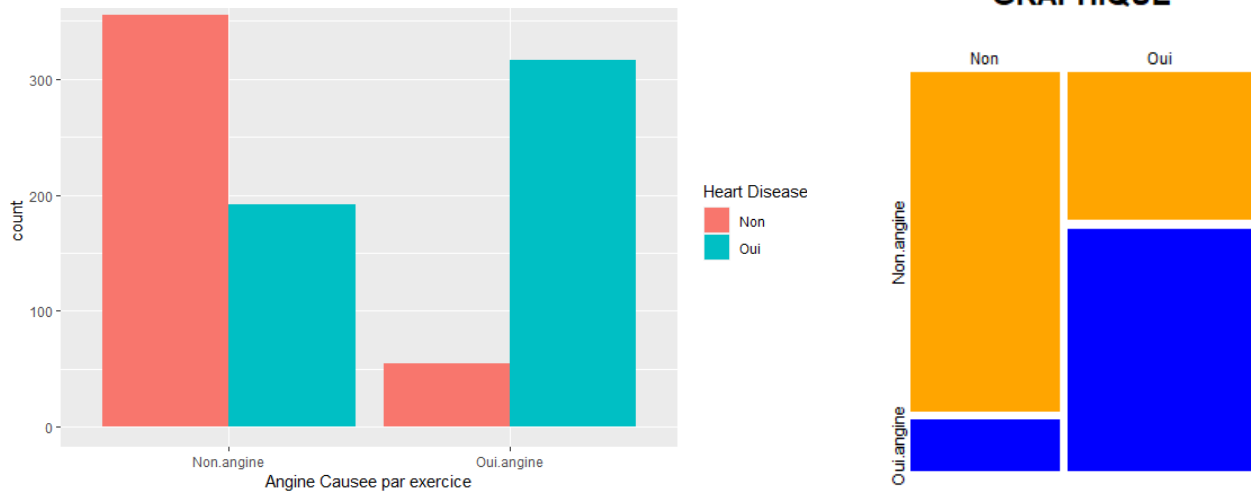
Le calcul des indicateurs de corrélation de Khi-deux et le V de cramer permettent de déterminer la vraie nature de la liaison entre les maladies du coeur et la forme du segment ST.

La valeur du V de cramer est 0.62 ce qui est **très élevé et est synonyme d'une forte liaison entre la forme du segment ST et les maladies cardiaques.**

6. Les maladies cardiaques sont-elles influencées par l'Angine causée par l'exercice ?

a. Analyse graphique

Débutons par visualiser la répartition des maladies du coeur par rapport l'Angine causée par l'exercice grâce à des diagrammes.



Ces graphiques nous donnent une idée de la liaison entre ces deux variables : *les individus présentant des angines après effort sont majoritairement plus sensibles aux maladies cardiaques*. Dans le but de clarifier cette probable relation, nous allons analyser le tableau de contingence, le tableau de fréquence et des profils et calculer les indicateurs de liaison qui sont le V de Cramer et le Khi-deux.

b. Analyse par des indicateurs

* Tableau statistique

Tableau De Contingence

	Non.angine	Oui.angine
Non	355	55
Oui	192	316

Tableau De Fréquence

	Non.angine	Oui.angine
Non	0.39	0.06
Oui	0.21	0.34

Tableau Des Profils Colonnes

	Non.angine	Oui.angine
Non	0.87	0.13
Oui	0.38	0.62

Tableau Des Profils Lignes

	Non.angine	Oui.angine
Non	0.65	0.15
Oui	0.35	0.85

L'analyse de ces deux tableaux montre que 34% de l'ensemble des patients ont une maladie du coeur et une angine de poitrine après avoir fait une activité physique. Aussi, 21% de l'ensemble des patients ont une maladie du coeur sans angine de poitrine après exercice.

En ce qui concerne la relation entre les maladies cardiaques et les douleurs dû à l'angine de poitrine suite à un effort physique, 62% des malades cardiaques ont une angine de poitrine après avoir fait un effort, c'est à dire que plus de la moitié des malades du coeur présentent des douleurs thoraciques après un effort physique. Alors que 38% des malades n'ont pas d'angine après exercices.

Par ailleurs, 85% des personnes ayant une angine suite à un exercice ont des maladies cardiaques.

* Indicateur de liaison

\$Khi_Deux
X-squared
222.2594
\$V_Cramer
[1] 0.494282
\$Khi2.P.value
[1] 2.907808e-50
\$Significativite_TestKhi2
[1] "liaison significative, les deux variables sont liées"
\$Fisher.P.value
[1] 3.464911e-54
\$Significativite_TestFisher
\$Intensite_liaison
[1] "liaison moyenne"

Le calcul des indicateurs de corrélation de Khi-deux et le V de cramer permettent déterminer la vraie nature de la liaison entre les maladies du coeur et l'Angine de poitrine.

Par conséquent, ***il existe une liaison significative d'intensité moyenne*** entre l'existence des **maladies cardiaques** et l'**Angine de poitrine après exercice**.

Conclusion

Cette dernière partie avait pour objectif de mettre à nu les relations de corrélation qui pourraient exister entre les maladies du coeur et certains facteurs.

Il en résulte que toutes les variables récoltées ont une influence sur la présence ou non de maladie cardiaque. Cependant, deux de nos variables dépendantes ont plus d'impact sur l'existence de la maladie chez ces individus.

Celle qui a la plus forte corrélation est la forme de la pente du segment ST, ensuite vient l'Angine de poitrine après exercice avec une intensité moyenne.

CONCLUSION GENERALE

En conclusion, notre étude a apporté des éclairages significatifs sur les facteurs contribuant au diagnostic des maladies cardiaques.

L'analyse univariée a révélé plusieurs variables individuelles ayant une association marquée avec la présence de maladies cardiaques, fournissant ainsi des indications cruciales pour l'identification des patients à risque. Cette première phase nous a permis de découvrir que globalement la majorité (79%) des patients sont de sexe masculin, ont un âge compris entre 44 ans et 63 ans et sont à 23% des diabétiques.

L'exploration des relations bivariées a enrichi notre compréhension en mettant en évidence des interactions complexes entre divers facteurs et les maladies cardiaques. Nous avons onze (11) variables dépendantes, tous ces paramètres ont une liaison plus en moins fortes avec les maladies du coeur. Par contre, notre étude a révélé deux principaux éléments qui impactent le plus la présence ou l'absence de maladie cardiaque. Le facteur ayant une forte intensité de liaison est la forme de la courbe du segment ST obtenue grâce à la mesure sur l'électrocardiogramme. Notre étude a montré que *78% de patients qui ont une dépression du segment ST ont une maladie cardiaque* et 20% de ceux qui ont une élévation du segment ST ont une maladie cardiaque.

Ces derniers résultats sont associés à des maladies telles que : l'infarctus du myocarde ou une ischémie myocardique (manque d'apport sanguin au muscle cardiaque). Ce qui confirme la forte intensité de la relation qui existe entre les deux variables.

Le second facteur influençant l'existence ou l'absence de maladie cardiaque est la présence de douleur thoracique juste après un effort.

L'analyse de la liaison montre ici que 85% des personnes ayant une angine suite à un exercice ont des maladies cardiaques. Et que 62% des malades cardiaques ont une angine de poitrine après avoir fait un effort.

Les résultats obtenus de notre étude portant sur les facteurs favorisant le diagnostic des maladies cardiaques sont conformes à la littérature médicale portant sur le sujet. Ils viennent donc en complément à la documentation existante. Cependant, ces considérations soulignent la nécessité d'une interprétation prudente des résultats et de futures recherches pour affiner nos conclusions.

Bibliographie

OMS. «Maladies cardiovasculaires.» *Organisation Mondiale de la Santé*. 17 Mai 2017.
[https://www.who.int/fr/news-room/fact-sheets/detail/cardiovascular-diseases-\(cvds\)](https://www.who.int/fr/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)) (accès le 11 30, 2023).

R, Source: Logiciel. s.d.

Source: Kaggle, nos calculs. «Heart Failure prediction Dataset.» Septembre 2021.