

## **Projet 9: STA101**

### **Analyse des données : Méthodes descriptives**

---

projet réalisé par : Bargaoui Emna

# Tables des matières

<b>Introduction</b>	<b>3</b>
<b>Analyse statistique élémentaire:</b>	<b>4</b>
Présentation et pré-traitement du jeu de donnée:	4
Analyse Univariée :	6
Analyse Bivariée:	10
<b>Analyse des correspondances multiples (ACM):</b>	<b>13</b>
Plan factoriel principal	14
Plan factoriel 3-4	18
<b>La classification non supervisée</b>	<b>21</b>
Choix du nombre de classes	22
Description des classes à partir des variables	22
<b>Conclusion</b>	<b>28</b>
<b>Annexes</b>	<b>30</b>

## Introduction

Les maladies cardiovasculaires sont la principale cause de décès dans le monde et surtout aux Etats-Unis chez les hommes comme chez les femmes.

Selon une étude statistique d'ici 2023 plus de 23 millions de personnes mourront en raison de problèmes liés au cœur.

Le développement des technologies et l'accès à l'information permet une meilleure prévention et des diagnostics plus élaborés pour ces maladies . On a des bases de données médicales plus exhaustives et complètes , c'est notamment ce que nous propose le "BRFSS"(un système de surveillance des facteurs et de risque comportementaux) établi en 1984 aux Etats-Unis .

Dirigé par plusieurs départements de santé , il consiste en une série d'enquêtes téléphoniques transversales menées sur des sujets tels que le mode de vie , la diététique ,les antécédents médicaux .

Notre sujet d'étude portera sur une base de données extraite de ce système , constituée principalement par certains facteurs de risques qui peuvent entraîner ou favoriser les maladies cardiovasculaires.

On veut pouvoir qualifier les profils de personnes atteintes de maladies cardiovasculaires sur la base de différents facteurs de risques tels que des antécédents médicaux et le mode de vie , mais aussi sur les conditions physiques et mentales .

Nous voulons identifier de possibles similitudes ou disparités entre ces malades ,on veut aussi identifier un lien entre des facteurs non contrôlables tels que l'âge ,le sex et la race et le reste des facteurs de risques cardiovasculaire et autres .

Pour mener à bien notre étude ,nous allons dans un premier temps décrire plus en détail le jeu de données et effectuer une analyse statistique univariée et bivariée.Puis nous réaliserons une analyse des correspondances multiples (ACM) suivie d'une méthode de classification non supervisée.

## I. Analyse statistique élémentaire:

### A. Présentation et pré-traitement du jeu de donnée:

Le jeu de données original contient à peu près 319.000 individus et 18 variables .

Les enquêtes se limitent au territoire américain ,notons aussi qu'il ya différents types de maladie du cœur , ici ce sont les maladies coronariennes (CHD) et les infarctus du myocarde qui sont retenus implicitement dans l'enquête.

Nous ne considérerons donc que les personnes atteintes de maladies cardiovasculaires, donc la variable "HeartDisease" ne sera pas prise en compte et on ne sélectionne que les lignes renvoyant "True" a celle-ci , ce sont eux nos individus statistiques.

Ceci nous laisse avec 27.000 individus , nous allons procéder à la sélection d'une sous-population tirée aléatoirement .Nous avons plus d'une dizaine de variables et les individus sont anonymes , il nous faudrait une taille raisonnable pour notre jeu de données , au moins 300 individus seraient convenables . Nous avons choisi 500 pour avoir des résultats plus stables et robustes à nos analyses statistiques en vérifiant qu'il y a au moins 10 observations par variables (par modalités ) pour éviter tout problème de calcul .

#### Description des variables

Antécédents médicaux /conditions médicales	<b>BMI</b>	Indice de masse corporel	Quantitative
Antécédents médicaux /conditions médicales	<b>Stroke</b>	Avez-vous déjà fait un AVC ? Vrai/Faux	Qualitative
Antécédents médicaux /conditions médicales	<b>Diabetic</b>	Aviez /Avez vous le diabète ? Vrai/Faux	Qualitative
Antécédents médicaux /conditions médicales	<b>KidneyDisease</b>	Aviez /Avez vous une maladie des reins(hors calcul reinaux et infections urinaires ) ? Vrai/Faux	Qualitative
Antécédents médicaux /conditions médicales	<b>SkinCancer</b>	Aviez /Avez vous un cancer de la peau ? Vrai/Faux	Qualitative
Antécédents médicaux	<b>Asthma</b>	Aviez /Avez vous de l'asthme ? Vrai/Faux	Qualitative

/conditions médicales			
mode de vie	<b>Smoking</b>	Avez-vous fumé au moins 5 paquets de cigarettes dans votre vie? Vrai/Faux	Qualitative
mode de vie	<b>AlcoholDrinking</b>	Êtes- vous un grand consommateur d'alcool ? (Hommes : 14 verres par semaine Femmes: 7 verres par semaine ) Vrai/Faux	Qualitative
mode de vie	<b>PhysicalActivities</b>	Si une activité physique est pratiquée en dehors du boulot (courir,faire du golf,jardiner,marcher..) Vrai/Faux	Qualitative
mode de vie	<b>SleepTime</b>	En moyenne combien d'heures dormez-vous dans une journée ?	Quantitative
Conditions physiques/mentales	<b>PhysicalHealth</b>	Pendant combien de jours durant les 30 jours passés votre état physique (limitations fonctionnelles,douleur, blessures,consommation excessive de médicaments,dépendance à autrui pour certaines tâches (ex ménagère )) n'était pas bien ?	Quantitative
Conditions physiques/mentales	<b>MentalHealth</b>	Pendant combien de jours durant les 30 jours passés votre état mental (dépression,anxiété,stress,problèmes émotionnelles) n'était pas bien ?	Quantitative
Conditions physiques/mentales	<b>Diffwalking</b>	Avez-vous des difficultés à marcher ou à monter des escaliers ? Vrai/Faux	Qualitative
Conditions physiques/mentales	<b>GenHealth</b>	Que pouvez-vous dire de votre santé en générale ? Excellent/Very good/Good/Fair/Poor	Qualitative
Autre facteurs non contrôlables	<b>AgeCategory</b>	13 modalités 18-24/25-29/30-34/35-39/40-44/45-49/50-54/55-59/60-64/65-69/70-74/75-79/80 or older	Qualitative

Autre facteurs non contrôlables	<b>Sex</b>	Femme ou Homme	Qualitative
Autre facteurs non contrôlables	<b>Race</b>	6 modalités Black/White/Asian/Hispanic/Other /American native-Indian	Qualitative

#### Autres précisions:

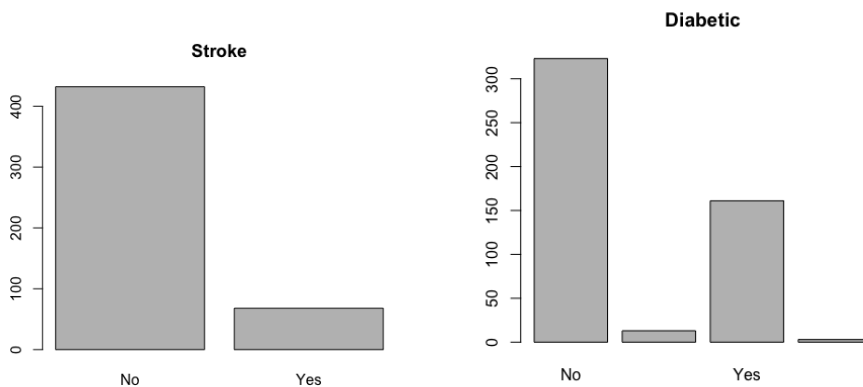
GenHealth: Cette variable fait référence à l'état de santé auto évalué qui est un outil efficace pour permettre un dépistage de groupe de personnes à risques de mortalité soudaine et en particulier chez les personnes âgées. Les sujets évaluent leur état de santé en se comparant aux autres personnes de la même tranche d'âge. Cette auto-évaluation se base sur des critères tels que la capacité à se rétablir d'une maladie, la fréquence d'utilisation des services médicaux, le bien être générale, de même que le statut socio-économique, la qualité de vie.

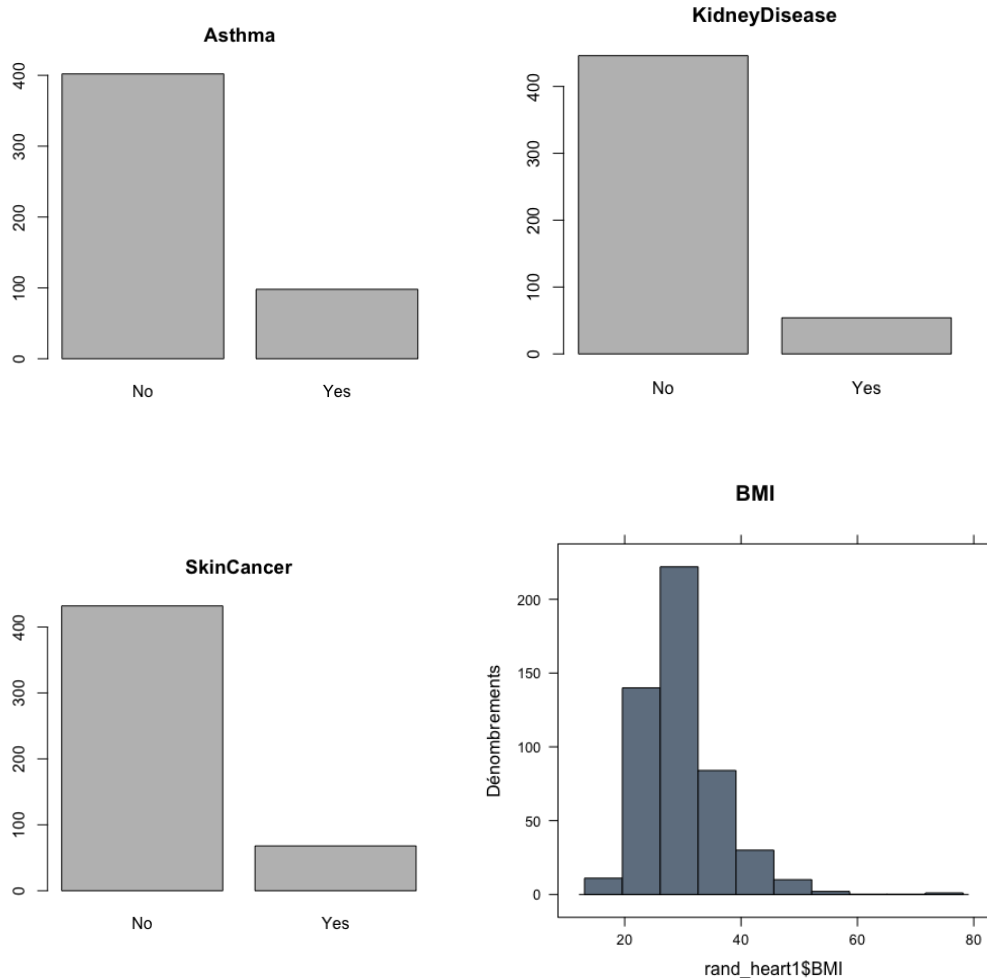
MentalHealth/PhysicalHealth: Des mesures souvent utilisées pour des personnes souffrant de maladies chroniques ici entre autres celles liées au cœur et qui nous donnent des informations sur la santé (mental/physique) récente de l'individu.

#### B. Analyse Univariée :

Tous les tableaux de fréquences pour les variables qualitatives sont en annexe 1

- Les antécédents médicaux

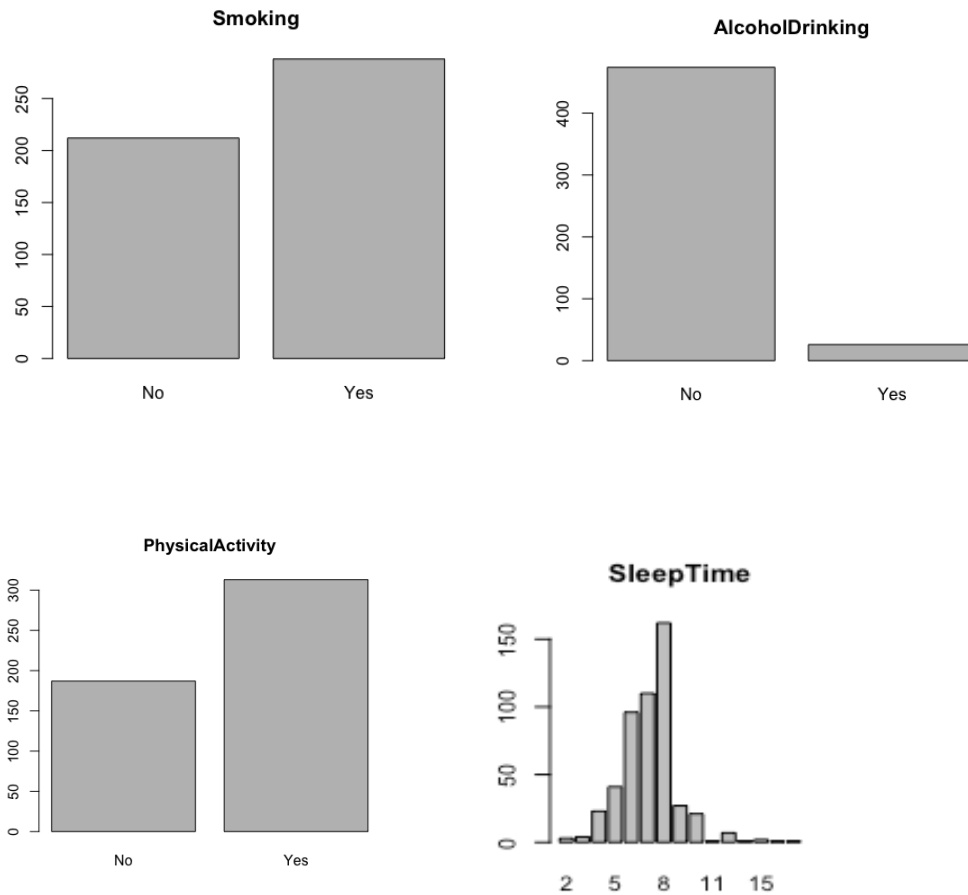




**BMI:** Min. 1st Qu. Median Mean 3rd Qu. Max.  
 15.50 25.38 28.70 29.62 32.75 75.77

- Stroke : Une grande majorité des malades n'ont jamais fait d'AVC
- Diabetic : 64% des malades n'ont pas d'antécédents de diabète contre 35% qui en ont. On remarque une très faible proportion voir inexistante pour les femmes l'ayant eu pendant leur grossesse et ceux qui ont eu un pre-diabète.
- Asthma : Plus des trois-quart ne sont pas asthmatiques (ou ne l'ont jamais été)
- KidneyDisease: Une grande majorité n'ont pas ou n'ont pas eu de maladie des reins.
- SkinCancer: Seulement une faible proportion ont eu un cancer de peau
- BMI : On observe une distribution asymétrique à droite, la moyenne est légèrement supérieure à la médiane et un test de Shapiro-Wilk nous confirme la non normalité (cf annexe 2). La moitié des individus ont un indice de masse corporelle entre 22.66 et 32.59. Selon l'OMS un IMC normal devrait se situer entre 18.5 et 24.9. Donc une majorité sont en excès de poids voir souffrant d'obésité. De même d'après la boxplot (cf annexe 3) on remarque des valeurs particulières mais d'après les données ce ne sont pas des erreurs techniques ou des valeurs improbables donc elles ne seront pas retirées. En effet un BMI supérieur à 40 signifie une obésité morbide ce qui n'est pas anodin aux Etats-Unis tandis que les quelques valeurs inférieures à 15-16 correspondent à des formes de sous-poids extrêmes.

- Mode de vie



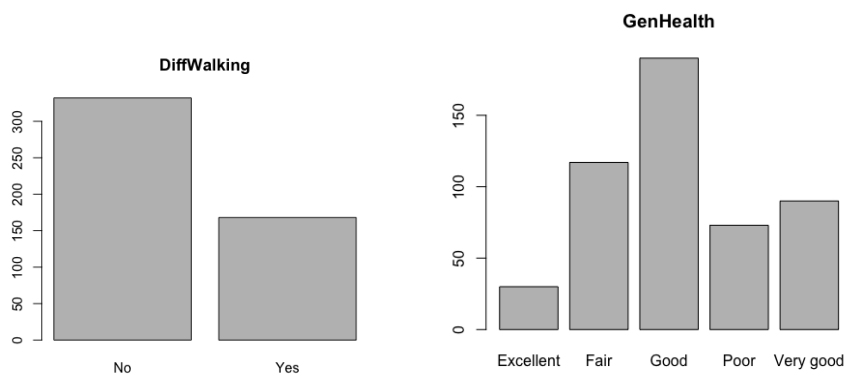
-Smoking: Un peu plus de la moitié sont des fumeurs

-AlcoholDrinking: Une très grande majorité ne sont pas de grands consommateurs d'alcool

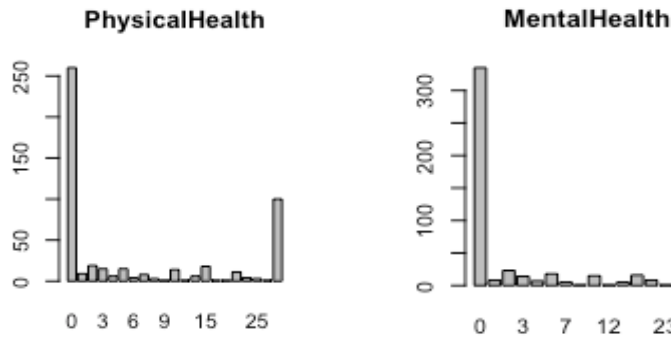
-PhysicalActivities: Pratiquement les deux-tiers font une activité physique en dehors du travail.

-SleepTime: On voit qu'une grande partie des individus dorment entre 5 et 9 heures. Une très faible partie 4 heures ou à l'extrême 10 heures. Les modalités en dessous de 4 heures ou supérieures à 10 heures se font très rares (sous-représentées).

- Conditions physiques et mentales







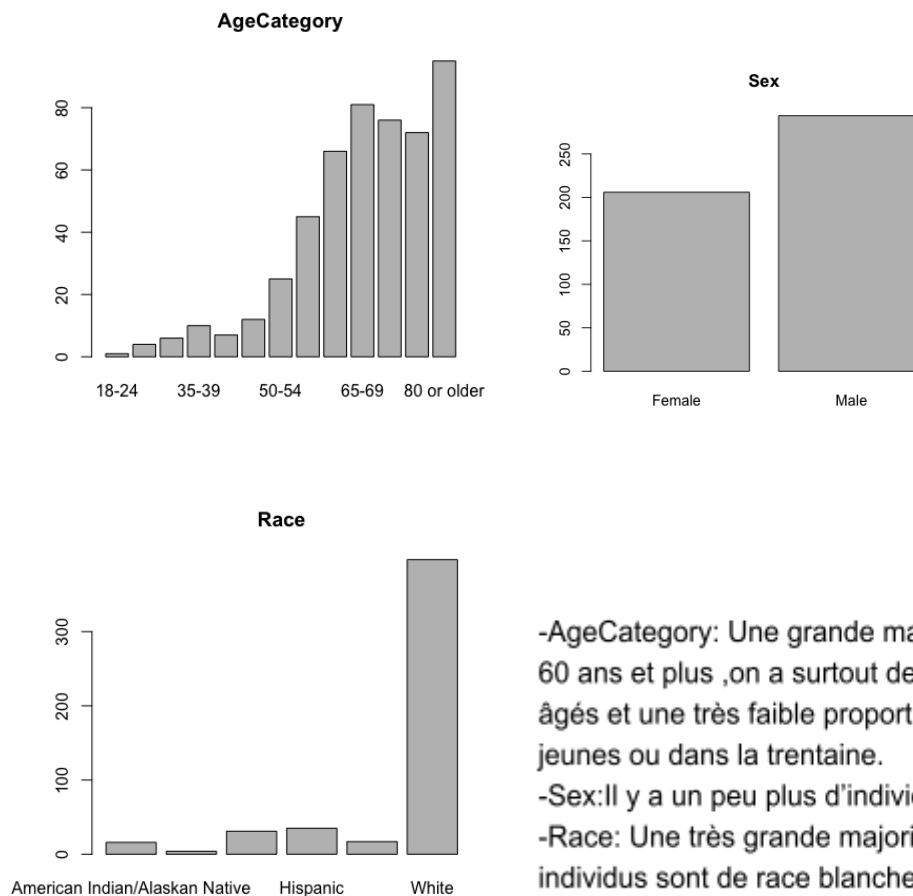
-DiffWalking: Une grande part non négligeable ont des difficultés à marcher ou monter des escaliers .

-GenHealth: La plupart pensent avoir une santé normale voir légèrement en dessous.

-PhysicalHealth: On remarque que la plupart des individus vont bien ,pas de blessure ,ni de limitation fonctionnelle,( puisque la classe 0 signifie que pendant 0 jour l'état physique de la personne n'allait pas bien).Une proportion non négligeable dit ne pas aller bien pendant 30 jours.

-MentalHealth: Comme pour la variable PhysicalHealth on observe deux pics aux extrémités (classe 0 et 30) .Une faible proportion dit ne pas allez bien mentalement pendant le mois passé .

- Autres facteurs "non contrôlables"



-AgeCategory: Une grande majorité ont entre 60 ans et plus ,on a surtout des individus âgés et une très faible proportion d'individus jeunes ou dans la trentaine.

-Sex: Il y a un peu plus d'individus hommes.

-Race: Une très grande majorité des individus sont de race blanche et une très faible est asiatique ou amérindienne.

Conclusion de l'analyse univariée:

A l'issue de cette analyse statistique , on a pu constater que certaines variables qualitatives possèdent une ou plusieurs modalités rares qui peuvent avoir un impact lors de l'ACM.Nous allons donc les regrouper de manière à conserver le plus fidèlement possible le sens des variables .

Pour la variable Diabetic on regroupe "yes" et "yes (during pregnancy)" et "no" avec "no borderline".Pour la variable AgeCategory les modalités en dessous de 50 ans sont sous représentées nous les avons regroupés ensemble sans que cela impacte le sens de la variable . Les autres modalités ont été regroupées en deux classes pour distinguer les personnes âgées ( plus de 65 ans ) d'après l'INSEE. La variable Race se compose de deux modalités "white" et "non white" qui regroupent le reste .

On constate également que pour les variables quantitatives SleepTime,PhysicalHealth et MentalHealth ,une très grande majorité des classes sont sous représentées , de même la majorité de nos variables sont qualitatives et nous souhaitons réaliser une ACM pour l'analyse factorielle donc nous allons requalifier toutes nos variables quantitatives y compris la variable BMI.Ajoutons que la discrétisation de l'indice de masse corporelle en plusieurs modalités facilitera sa compréhension .

De manière générale pour le découpage en classe nous procédons avec la méthode du seuil observé , en nous aidant des graphiques de distributions et en choisissant nos nouvelles classes de manière manuelle.

Pour MentalHealth et PhysicalHealth on aura trois modalités , "Aucun jour" qui est constituée du premier pic (classe 0) , "Pendant un mois" pour le dernier pic (classe 30) et "Nombreux jours" pour le reste entre ces deux pics .La variable SleepTime sera elle aussi découpée en trois modalités, [2-5] (les heures de sommeil entre 2 et 5 heures ) ,[6-8] (les heures de sommeil entre 6 et 8 heures ) et [9-plus] (les heures de sommeil de plus de 9 heures).

Enfin pour la variable BMI le découpage se fait en nous aidant de l'histogramme des indicateurs de positions et selon les indicateurs IMC de l'OMS. Nous avons quatre modalités , "sous-poids" ([15.5,18.5] ) , "normal"([18.5,25] ) , "surpoids" ((25,30] ) et "obésité/obésité morbide" ( > 30).(cf annexe 4 nouveaux diagrammes en barres)

### C. Analyse Bivariée:

Pour l'analyse bivariée nous allons nous concentrer sur les variables liées aux facteurs non contrôlables et les variables liées aux conditions médicales . L'analyse factorielle et la classification nous apporteront une vision multidimensionnelle avec les autres variables .

Nous allons étudier s'il existe un lien significatif avec un test de khi deux ou un test exact de Fisher lorsque les conditions du khi deux ne sont pas remplies( les résultats seront en annexe 5).On pourra aussi observer les coefficients d'associations comme le V de Cramer pour déterminer si nous avons des associations fortes ou non en cas de liaison .

- AgeCategory-Asthma

D'après le test du khi deux il y a un lien significatif ( $p\text{-value} = 0.01$ ) au risque de 5 % entre ces deux variables et d'après le coefficient de Cramer (0.13 ) nous avons une association

d'intensité moyenne.

-AgeCategory-Asthmatic:			
	No	Yes	Total
moins de 50 ans	67.5	32.5	100.0
50-65	75.7	24.3	100.0
65 or more	84.0	16.0	100.0
Ensemble	80.4	19.6	100.0

D'après le tableau profil-ligne ci-dessus parmi les 50 ans ,32% d'entre eux font (ont fait) de l'asthme alors que ceux ci ne représentent que 19 % de la population total .Or les plus de 65 ans 16% d'entre eux le sont(ou étaient).

- AgeCategory-SkinCancer

D'après le test exact de fisher il ya un lien significatif entre les deux (p-value=0.01) au risque de 5 % .Le coefficient de cramer est de 0.13, on a une association d'intensité moyenne. D'après le tableau (cf annexe 6) 13.6% de la population ayant une maladie cardiovasculaire ont eu un cancer de la peau et cette proportion est sensiblement plus élevée chez les plus de 65 ans avec 16.7%.

- AgeCategory-BMI

D'après le test exacte de fisher il ya un lien significatif entre les deux (p-value=0.004) au risque de 5 % .Le coefficient de cramer est de 0.14 on a une association d'intensité moyenne. D'après le tableau (cf annexe 6) les individus entre 50 et 65 ans 51% d'entre eux sont obese ou extrêmement obese ,alors que ces derniers représentent 39% de la population . La proportion étant supérieure à la proportion moyenne dans la population totale , on peut bien voir qu'il y a un bien un lien intéressant entre ces deux modalités .

- AgeCategory-Diabetic AgeCategory-KidneyDisease AgeCategory-Stroke

Le test du khi deux et de fisher pour ces couples nous renvoie respectivement 0.2 0.3(fisher) et 0.07 pour la valeur de p-value donc on ne peut rejeter l'hypothèse d'indépendance et on ne peut conclure s' il existe une relation fiable entre les deux .

- Race-Diabetic

D'après le test du khi deux il y a un lien significatif (p-value = 0.02) au risque de 5 % entre ces deux variables et d'après le coefficient de cramer (0.1 ) nous avons une association d'intensité faible. D'après le profil ligne (annexe 6) parmi la catégorie"White" 30% d'entre eux ont eu (ou l'ont) le diabète alors que ceux ci représentent 32.8% de la population totale.

- Race-Asthma

D'après le test du khi deux il y a un lien significatif (p-value = 5.524e-06) au risque de 5 % entre ces deux variables et d'après le coefficient de cramer (0.2 ) nous avons une association d'intensité forte. Le profil ligne (annexe 6 ) nous montre que 15% de la race "White" font (ou ont fait ) de l'asthme.

- Race-KidneyDisease

D'après le test du khi deux il y a un lien significatif (p-value =0.008) au risque de 5 % entre ces deux variables et d'après le coefficient de cramer (0.12 ) nous avons une association d'intensité moyenne. Le profil ligne (annexe 6) nous montre que 18% des autres races ont eu une maladie des reins et représentent 10.8% de la population totale.

- Race-BMI Race-SkinCancer Race-Stroke

Le test du khi deux et de fisher pour ces couples nous renvoie respectivement 0.4(fisher) ,0.07 et 0.6 pour la valeur de p-value donc on ne peut rejeter l'hypothèse d'indépendance et on ne peut conclure s' il existe une relation fiable entre les deux

- Sex-Asthma

D'après le test du khi deux il y a un lien significatif (p-value =0.036) au risque de 5 % entre ces deux variables et d'après le coefficient de cramer (0.098 ) nous avons une association

d'intensité moyenne. D'après le profil ligne , parmi les femmes 24.3% d'entre elles font (ou ont fait) de l'asthme alors que ces derniers représentent 19.6% de la population. Leur proportion est supérieur à la proportion moyenne dans la population totale il ya bien un lien intéressant entre ces deux variables.

- Sex -Diabetic Sex-Stroke Sex-SkinCancer Sex-BMI Sex-KidneyDisease

Le test du khi deux et de fisher pour ces couples nous renvoie respectivement 1, 0.3 , 0.5, 0.3, et 0.1 pour la valeur de p-value donc on ne peut rejeter l'hypothèse d'indépendance et on ne peut conclure s' il existe une relation fiable entre les deux.

Conclusion analyse bivariée :

A l'issue de cette analyse bivariée les variables liées à des facteurs non contrôlables sont liées à d'autres issues des "conditions médicales/antécédents médicaux" notamment la variable Asthma qui semble liée aux trois (sex, age , race)

## II. Analyse des correspondances multiples (ACM):

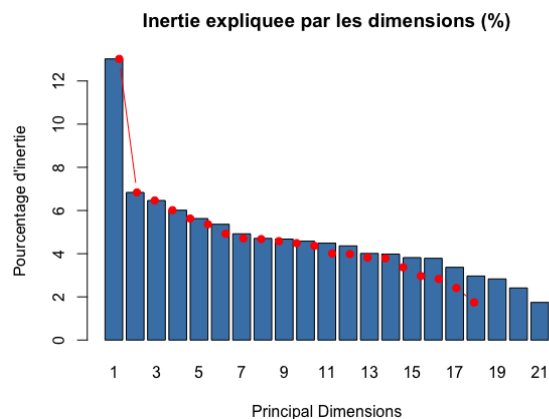
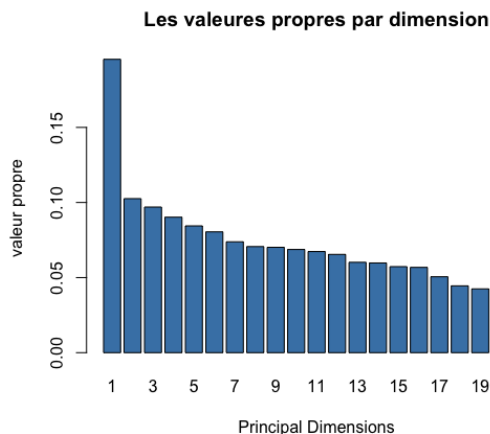
Nous allons maintenant effectuer une analyse factorielle sur nos données qui nous permettra entre autres d'obtenir une typologie de l'enquête basée sur différents facteurs comme les antécédents médicaux ,le mode de vie et les conditions physiques et mentales (globales ou récentes).

Donc les variables actives sont les 14 variables répertoriées dans les trois catégories mentionnées ci dessus. Nous voulons aussi identifier le lien entre les facteurs tels que le sex, l'âge et la race des individus et le reste , les variables AgeCategory , Race et Sex seront des variables illustratives.

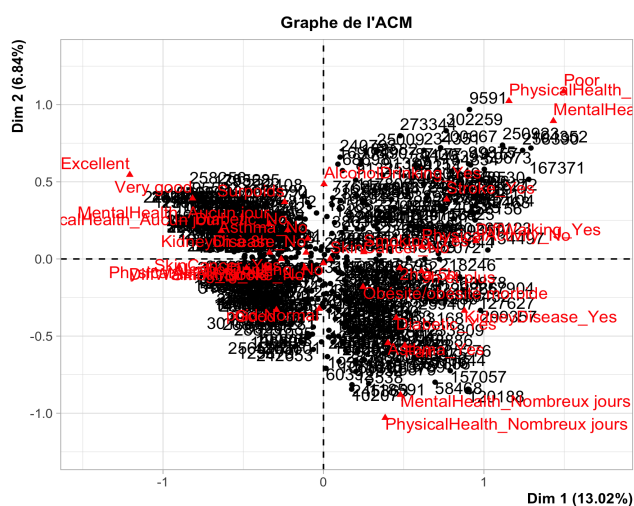
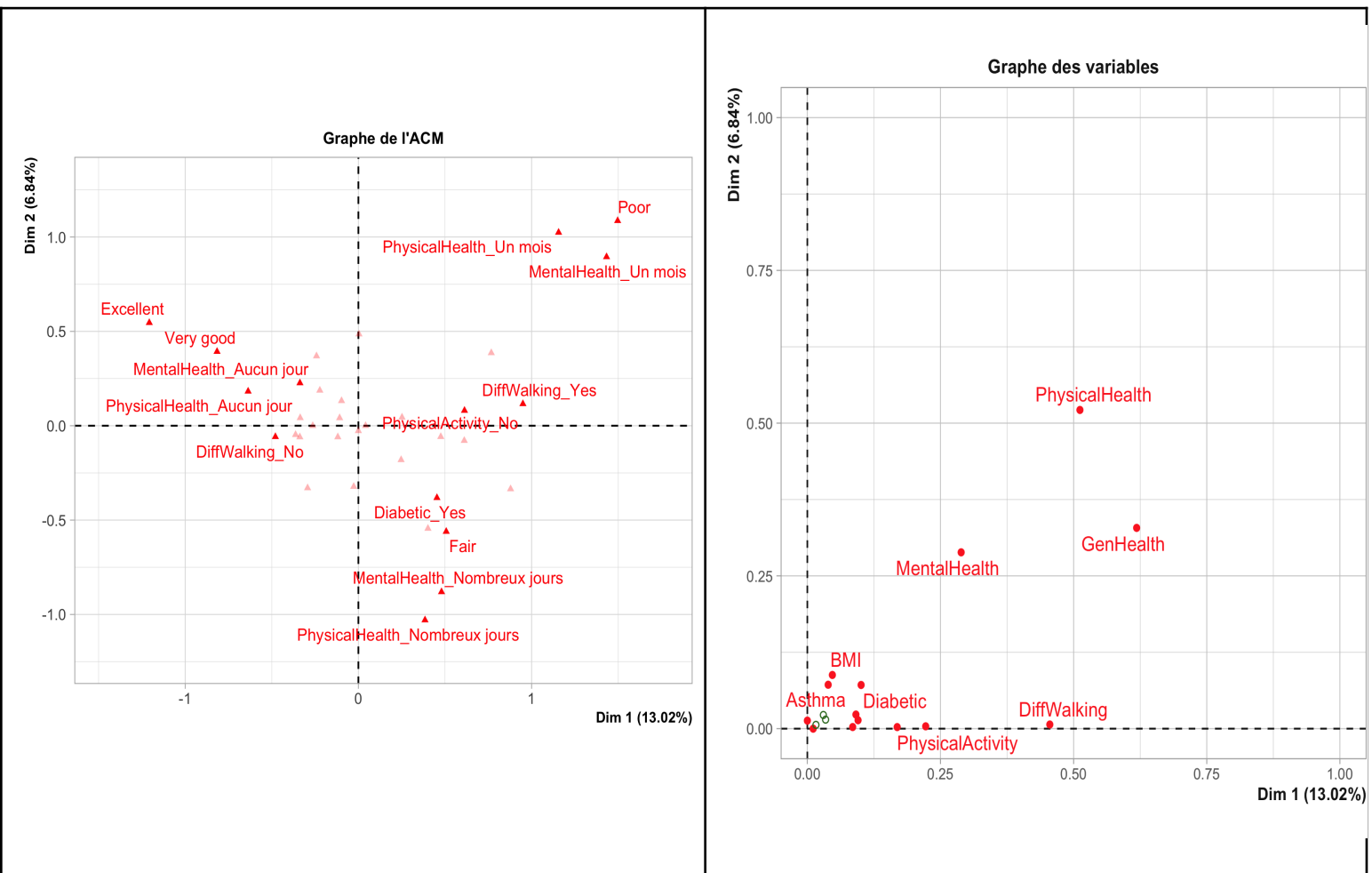
Notons que toutes nos variables actives sont bien qualitatives pour pouvoir se prêter à une ACM.

Nous avons d'abord effectué une première projection sur le premier plan factoriel et on a constaté que la modalité "sous-poid" était clairement isolée par rapport aux autres en raison de son très faible effectif observé lors de l'analyse univariée.Nous allons donc supprimer pour la suite de l'analyse les 8 individus ayant cette modalité (cf annexe 7).

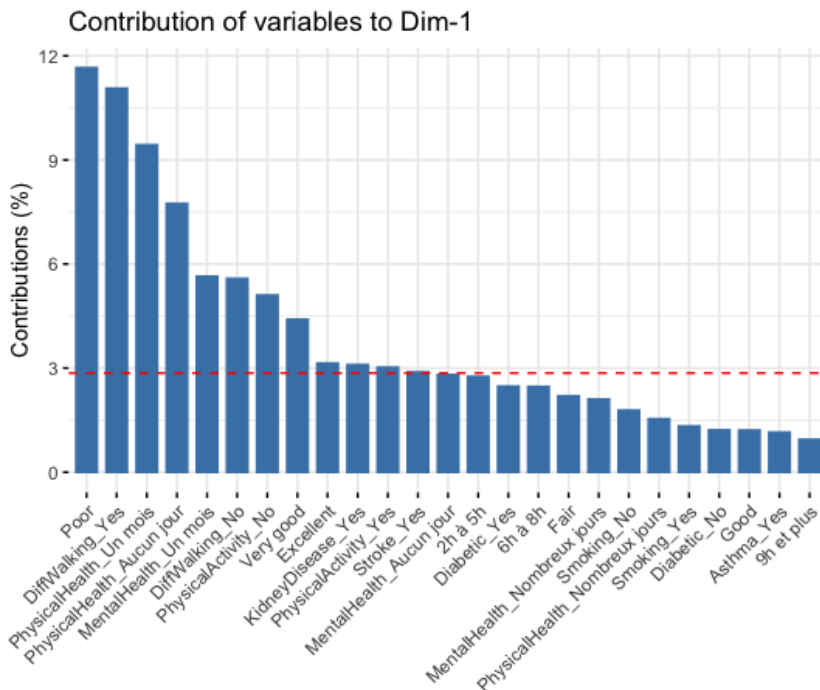
Nous allons commencer par décider combien d'axes nous allons explorer en utilisant la règle "à la Kaiser". On retiendra les axes dont la valeur propre est supérieur à  $(1/\text{nombre de variables actives})$  (0.071).Ceci nous suggère d'analyser jusqu'à 7 dimensions.On se limitera dans le cadre de ce projet aux 4 premiers qui résument 32% de l'information (annexe 10).



## A. Plan factoriel principal



Tout d'abord analysons de manière sommaire le nuage des individus puis nous le décrirons plus en détail par la suite avec l'interprétation des axes associés. Il a une forme particulière en parabole connu sous le nom "d'effet Guttman" qui rend compte d'une organisation hiérarchique des individus. Le premier axe est censé répartir les individus selon des modalités croissantes, une opposition de modalités extrêmes, et le deuxième axe est censé opposer des individus aux modalités moyennes avec des extrêmes.



Le graphe des variables nous permet de voir quelles variables sont le plus liées à cet axe , on a ici un groupe de trois variables PhysicalHealth MentalHealth et GenHealth puis ,PhysicalActivity et DiffWalking .La représentation des modalités et leur contribution va aider à avoir une interprétation plus fine.

Cette dimension met en avant des profils très distincts .

Nous avons une opposition entre des individus qui confirment avoir une mauvaise santé générale (hormis leurs problèmes cardiovasculaires),avec un état physique (blessures,injures majeures ) et mental (fatigue mentale ,dépression..) qui n'allait pas durant le mois passé (à droite de l'axe),donc en amont de l'enquête ,à ceux ayant une très bonne voir excellente santé générale et qui sur le plan physique et mental tout allait bien pendant tout le mois (à gauche de l'axe). La modalité "kidneyDisease\_Yes" contribue également à la construction de l'axe (droite) , on retrouve des individus atypiques sur cette modalité(cf analyse univariée , effectif faible)

De même nous avons d'un côté (à droite ) des individus qui ont des difficultés à marcher ou monter des escaliers ,des pentes ,qui ne pratiquent aucune activité physique en dehors de leur travail à ceux n'ayant aucun trouble de la marche et qui ont tendance à se dépenser physiquement hors boulot(à gauche).

La table des liaisons(cf annexe 11 ) des modalités de la dimension 1 illustre bien cette opposition ,ceux du côté droit ont une liaison positive et du côté gauche négatif.

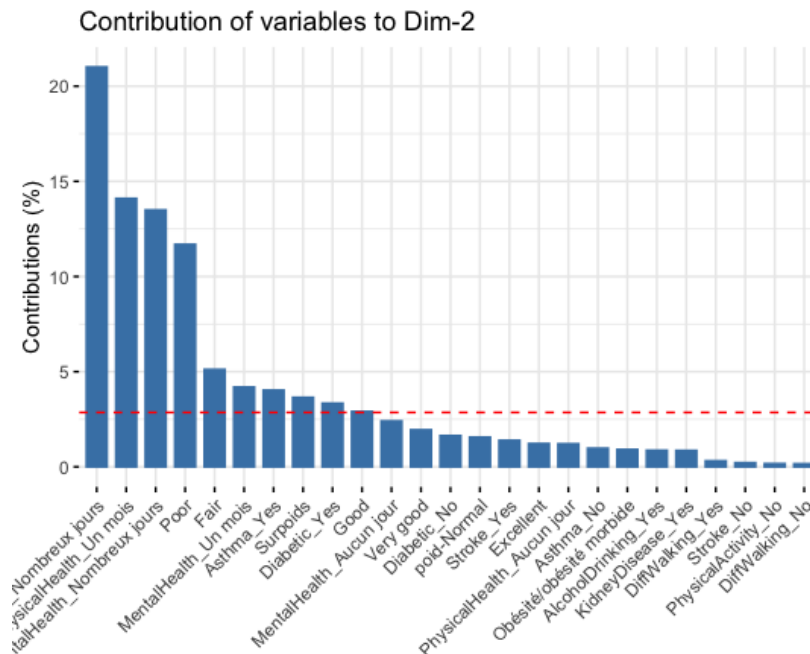
Pour appuyer notre interprétation nous revenons aux données brutes , nous sélectionnons deux individus aux positions extrêmes qui contribuent à la création de l'axe (cf annexe 8) .

L'individu "258256" avec une coordonnée négative sur l'axe 1 dont la perception sur sa santé générale est très positive tant sur l'aspect mental et physique, qui n'a aucune difficulté à se déplacer et qui peut se permettre une pratique physique s'oppose à "167371".

Celui-ci ne pratique aucune activité physique après le travail ou le week-end , et a une mobilité réduite . Il donne un retour très négatif sur son état de santé générale , de même il

affirme qu'il ne se sentait pas bien pendant tout le mois au niveau mental et physique .Enfin ,il est aussi atteint d'une maladie des reins .

- L'axe 2



Le graphe des variables nous permet de voir quelles variables sont le plus liées à cet axe , on a ici un groupe de trois variables PhysicalHealth MentalHealth et GenHealth comme pour l'axe 1.La représentation des modalités et de leur contribution vient affiner notre interprétation.

Cet axe oppose les individus(en bas) diabétiques,qui se sont plaints d'avoir pendant plusieurs jours des problèmes sur le plan mental comme physique (injuries,limitations fonctionnelles,douleurs..) à ceux pour qui ça a duré tout le mois passé (en haut).

Il oppose aussi les individus (en haut) affirmant avoir une mauvaise santé générale et en surpoids, à ceux pour lesquels celle-ci est plutôt moyenne voir bonne(en bas) .

Ici aussi on choisit des individus extrêmes qui ont contribué à la création de l'axe pour appuyer notre interprétation ( annexe 8).

La dimension 2 oppose l'individu "9591" (coordonnées fortement positive ) qui se perçoit en mauvaise santé , et qui durant tout le mois s'est dit ne pas être bien émotionnellement et physiquement à l'individu "120188"(coordonnées fortement négative ).Ce dernier a une santé générale moyenne et s'est aussi plaint de ne pas s'être senti bien mentalement et physiquement mais pour plusieurs jours ,et est diabetique.

Le principal plan factoriel nous a permis d'observer différents profils de personnes atteintes de maladies cardiovasculaires qui sont bien mis en avant par la forme parabolique du nuage des individus.

Le sommet de la parabole contient la majorité des individus aux modalités moyennes pour les variables GenHealth,PhysicalHealth,MentalHealth("plusieursjours", "good", "fair").

Il y a ceux qui se sentent bien ,rien d'alarmant en général,ceux qui pensent que leur qualité de vie , bien être ... est plutôt dans la moyenne.Rappelons que ces individus donnent leur

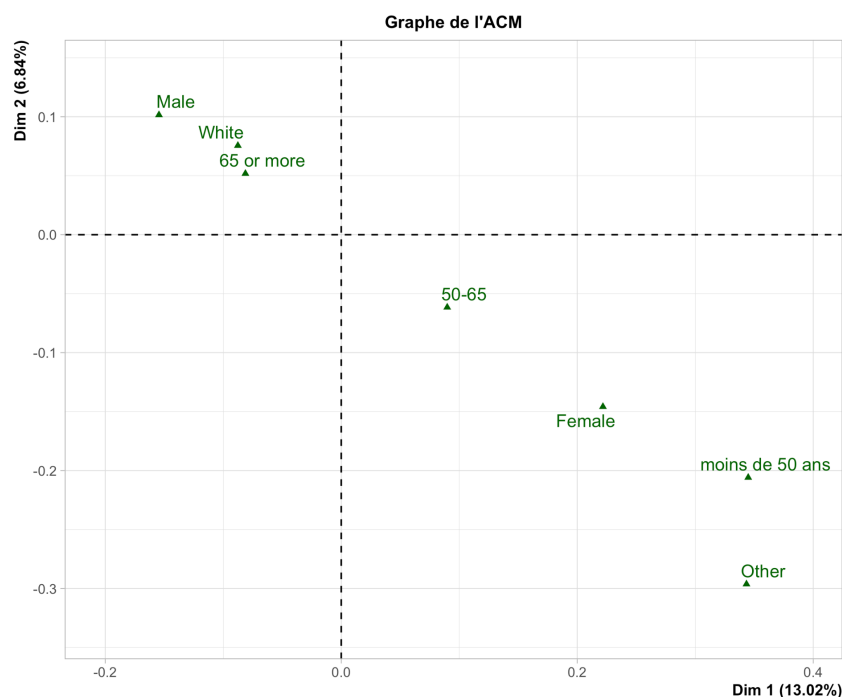


ressenti aussi par rapport à leur entourage , aux personnes de la même tranche d' âge qu'eux , à leur statut socio-économique..On retrouve aussi des individus vivant avec le diabète (ou l'étaient) et qui pourrait avoir fortement contribué à leurs maladies chroniques du coeur.

Aux extrémités on a une opposition marquée avec une majorité des individus possédant une santé précaire et ceux dont la santé se porte très bien malgré les maladies touchant le cœur ,et qui peuvent se dépenser physiquement .Plus précisément nous avons ceux qui pensent avoir une très bonne qualité de vie , les maladies mineures ou autres plus graves ne semblent pas les affectés sur leur bien être au quotidien.Ils pensent avoir une bonne capacité à se rétablir ou choisissent un comportement qui leur permet de maintenir une vie active dans leur foyer ou en dehors.Ces individus ne reportent aucune détresse émotionnelle, stress, ou une quelconque limitation physique , douleur physique qui aurait impacté négativement leur vie récemment.Leur état de santé récent est un bon indice aussi sur la qualité de leur santé globale et ils semblent vouloir maintenir celle ci puisqu'ils s'adonnent à une activité physique ce qui est fortement recommandé lorsqu'on est atteint d'une maladie chronique ( cardiovasculaire).

D'un autre côté il y a une majorité des individus qui en dehors des allers retours du travail ne pratiquent aucun sport ,ou toute forme de dépense physique .Ceci est peut être dû à une non volonté de leur part , ou au fait qu'ils ne peuvent bien se déplacer, malgré qu'il existe des sports adaptés à ce type de situation..Ce n'est donc pas particulièrement surprenant que ce type d'individu ait un très mauvais avis sur leur santé générale et même récente .Notons pour finir qu'on retrouve aussi des personnes avec la présence de maladies rénales qui peuvent être un des facteurs déclencheurs majeurs à leurs maladies cardiovasculaires.

Analysons maintenant les variables supplémentaires.

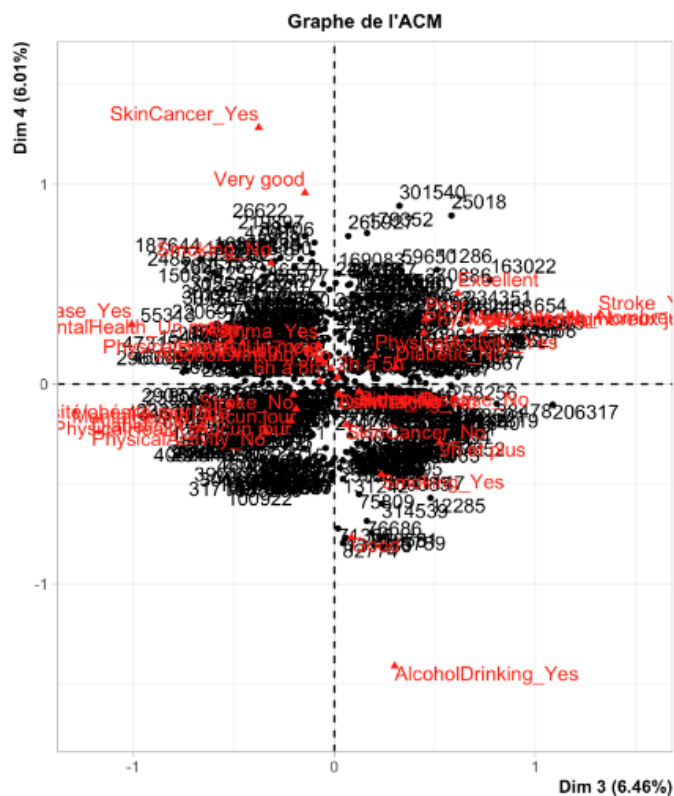


D'après le tableau de liaison entre les modalités et l'axe 1 et 2 (cf annexe 11 ) on peut observer que les variables qualitatives Sex et Race sont peu liées aux deux axes mais au vu du grand nombre des individus ces liaisons sont significatives. AgeCategory est liée à l'axe 1 et s'organise en ordre décroissant de gauche à droite.

On retrouve principalement des hommes atteints de maladies cardiovasculaires, blancs et vieux (plus de 65 ans ) corrélés aux individus qui pratiquent une activité physique dans la semaine ou en week end et qui n'ont pas de trouble de la marche .De même les individus de plus de 65 ans se perçoivent en très bonne santé , ceci n'est pas surprenant , selon plusieurs études les personnes plus avancées dans l'âge ont une impression plus favorable à leur santé, ils ont moins d'attentes .

Les femmes atteintes de maladies cardiovasculaires de moins de 50 ans non blanches sont corrélées aux individus qui ont du mal à marcher et qui ne pratiquent aucune dépense physique en dehors du travail. De même elles sont corrélées à deux groupes d'individus ceux disant avoir une santé médiocre et ceux dont la santé est correcte de manière générale malgré leur situation .

## B. Plan factoriel 3-4





Le graphe des variables ci-dessus nous montre que les variables BMI ,Diabetic et Stroke sont les plus liées à l'axe 3 .

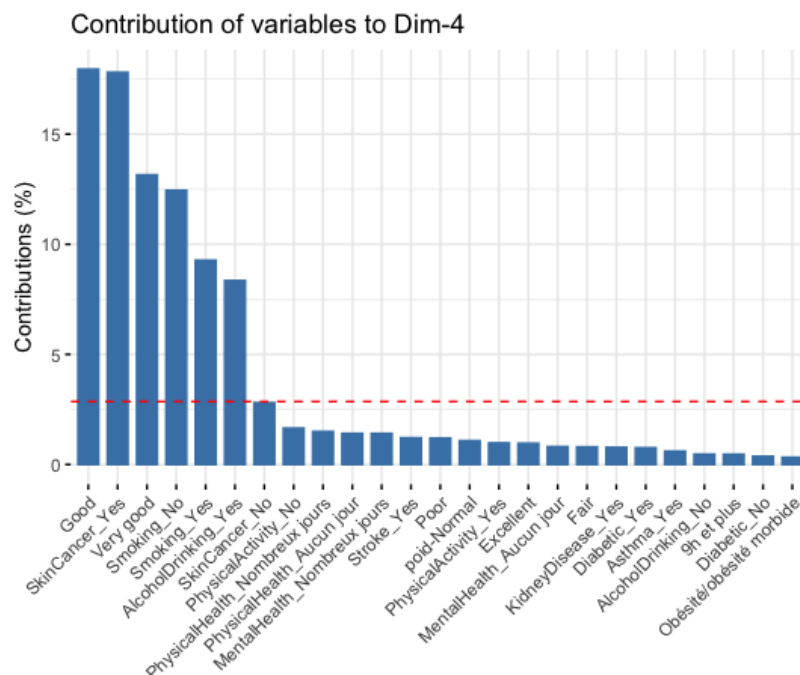
Il oppose des individus (côté droit) avec un poids normal sans antécédent de diabète et ayant déjà fait au cours de leur vie un AVC aux individus (cote gauche) atteints de maladies rénales, diabétiques (ou l'étaient) ,obèses (obèses extrêmes) ,ne pratiquant aucune activité physique et n'ayant jamais fait d'AVC. Ce qui n'est pas anodin car le diabète et l'obésité sont deux maladies très liées , en effet l'obésité est le principal facteur de risque du diabète.

Les variables "Smoking\_No" (cote gauche) et "MentalHealth\_Nombreux jours" "PhysicalHealth\_Nombreux jours " (cote droit) contribuent aussi à la création de cet axe.

On retrouve sur cet axe des individus atypiques sur la variable "KidneyDisease" et "Stroke " (effectif faible pour "Stroke\_yes" et "KidneyDisease\_Yes").

Revenons au données brutes (annexe 9 ),la dimension 3 oppose l'individu "206317" qui a un poids normal ,non diabetique et ayant déjà eu un AVC à l'individu "47315".Celui ci a un poids extrême avec des antécédents de diabètes ( ou toujours) mais n'ayant jamais eu d'attaque cérébrale (les autres modalités pour ces individus sont aussi respectées).

- L'axe 4



Ici aussi en nous aidant des modalités qui contribuent le plus , on peut dire que l'axe 4 oppose des individus qui avaient déjà eu un cancer de la peau et ne sont pas de grands fumeurs (du moins pas beaucoup)(en haut) à des individus qui n'ont jamais eu la mélanome et qui fument (en bas)

De même, il oppose ceux qui affirment avoir une très bonne santé globalement (en haut) à ceux pour qui elle semble bien (en bas).

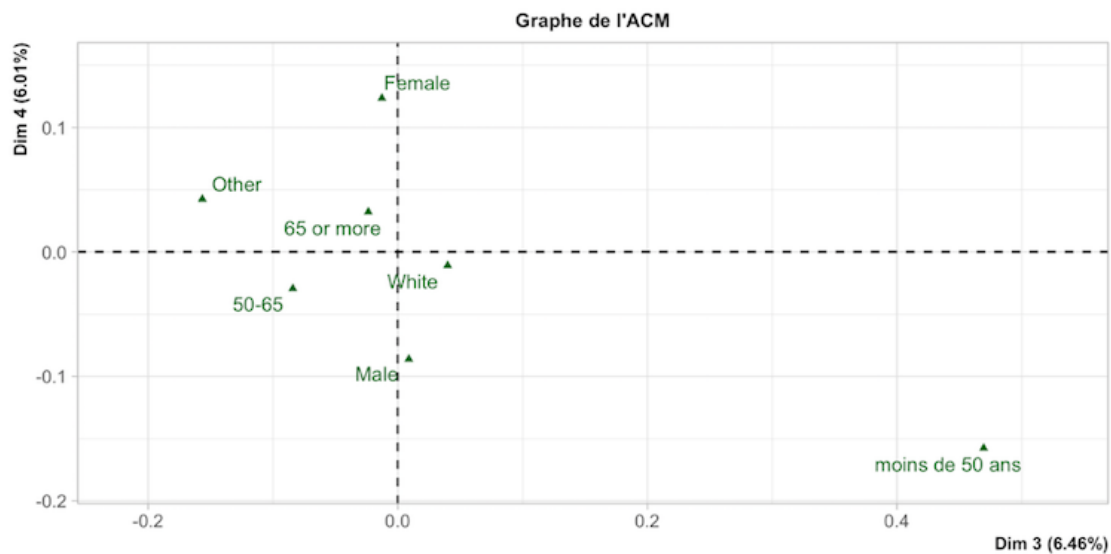
La modalité "AlcoholDrinking\_Yes" contribue aussi à la création de l'axe ,or lors de l'analyse univariée nous avons vu que son effectif est très faible est peut donc expliquer cette contribution donc nous décidons de ne pas l'inclure dans notre interprétation.

L'individu "26622" qui n'est pas un grand fumeur ,a déjà eu de mélanome et une très bonne santé au moment de l'enquête s'oppose à l'individu "12285". Celui-ci fume beaucoup mais se dit avec une bonne santé et n'a jamais eu de cancer de la peau (annexe 9) .

Pour la dimension 3-4 , le nuage d'individus a aussi une forme de parabole (effet Guttman) . Au sommet de la parabole on trouve la majorité des individus qui fument mais qui se disent en bonne santé en dépit de leur hygiène de vie .

La parabole marque une opposition évidente entre les individus sur leur corpulence et leur statut diabétique .On a les individus souvent obèses (ou extrêmement obèses ) et qui sont diabétiques surement aussi en raison de leur poids , et les individus qui maintiennent un poids normal et non diabétiques.

Analysons maintenant les variables supplémentaires.



D'après le tableau de liaisons entre les modalités et l'axe 3 et 4 (cf annexe 11) on peut observer que la variable qualitative Sex s'organise selon l'axe 4 et AgeCategory selon l'axe 3 .

Donc en haut on retrouve des individus femmes qui ont déjà eu un cancer de la peau et qui ne sont pas de grandes fumeuses .Les hommes sont corrélés aux côtés des individus n'ayant jamais eu un cancer de peau et qui sont qualifiés de grands fumeurs .

De même les individus de moins de 50 ans sont corrélés au 3ème axe du côté des individus avec une morphologie normale non diabétique mais ayant déjà fait au cours de leur vie un AVC.

### III. La classification non supervisée

La classification non supervisée va nous aider à identifier des groupes d'individus au profil similaire.

Nous allons effectuer une CAH sur nos données catégorielles et pour cela il faut appliquer celle-ci sur le résultat d'une ACM.

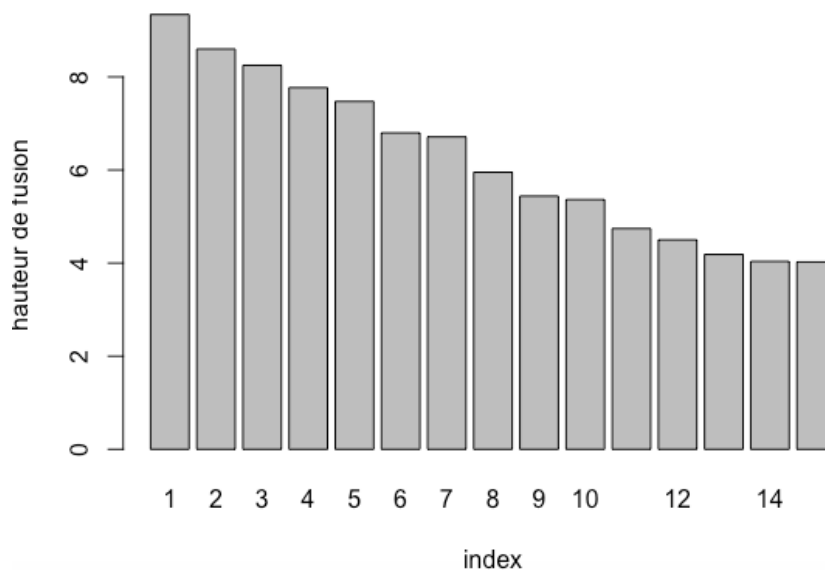
On conserve les 17 premiers axes de l'ACM qui capturent 90% de l'information ,le reste étant considéré comme du bruit.

Le critère d'agrégation sera celui de Ward qui permet de minimiser l'augmentation de l'inertie intraclasse, et la distance est la distance euclidienne usuelle.

A l'issue de la CAH on procédera à une consolidation avec la méthode des Kmeans pour sélectionner la meilleure partition en termes d'inertie.

#### A. Choix du nombre de classes

Le graphique des gains d'inertie intra classe nous montre une cassure entre la 6ème et la 5ème classe, on en étudiera donc 6 pour la partition.



La partition finale des classes est obtenue après l'application de la consolidation par Kmeans.(en annexe 13 on retrouve les graphes de la classifications sur les différents plans factorielles)

#### B. Description des classes à partir des variables

1	2	3	4	5	6
24	88	236	30	29	85

effectif des classes

On observe une classification déséquilibrée, la classe 3 est celle qui comporte le plus d'individus soit pratiquement 50% de l'ensemble des individus. La classe 1 est celle qui en contient le moins, seulement 24 juste derrière la classe 5.

	p.value	df
GenHealth	1.701273e-232	20
AlcoholDrinking	4.531949e-96	5
SleepTime	6.231232e-91	10
PhysicalHealth	6.508271e-46	10
DiffWalking	1.363175e-25	5
MentalHealth	3.126206e-13	10
PhysicalActivity	2.973608e-10	5
Smoking	5.359570e-10	5
SkinCancer	3.808101e-03	5
KidneyDisease	4.680852e-03	5
Diabetic	5.661646e-03	5
Stroke	1.261009e-02	5
BMI	1.272267e-02	10
AgeCategory	3.029989e-02	10

#### Classe 1

	Cla/Mod	Mod/Cla	Global	p.value	v.test
GenHealth=Excellent	100.0000000	100.0000000	5.894309	1.758165e-47	14.474414
PhysicalHealth=PhysicalHealth_Aucun jour	10.1167315	89.655172	52.235772	1.415752e-05	4.341404
DiffWalking=DiffWalking_No	8.5626911	96.551724	66.463415	8.320653e-05	3.934972
PhysicalActivity=PhysicalActivity_Yes	8.4142395	89.655172	62.804878	1.129509e-03	3.256108
SleepTime=6h à 8h	8.1632653	82.758621	59.756098	7.669733e-03	2.666273
MentalHealth=MentalHealth_Aucun jour	7.5301205	86.206897	67.479675	2.204530e-02	2.289586
Smoking=Smoking_No	8.5714286	62.068966	42.682927	3.343533e-02	2.126817
BMI=poid-Normal	10.3773585	37.931034	21.544715	3.926050e-02	2.061446
Smoking=Smoking_Yes	3.9007092	37.931034	57.317073	3.343533e-02	-2.126817
BMI=Obésité/obésité morbide	3.0456853	20.689655	40.040650	2.653664e-02	-2.218267
PhysicalHealth=PhysicalHealth_Nombreux jours	2.1428571	10.344828	28.455285	1.989778e-02	-2.328270
SleepTime=2h à 5h	2.4390244	13.793103	33.333333	1.743411e-02	-2.377422
GenHealth=Poor	0.0000000	0.0000000	14.227642	1.011618e-02	-2.571833
GenHealth=Very good	0.0000000	0.0000000	18.089431	2.535975e-03	-3.019016
PhysicalHealth=PhysicalHealth_Un mois	0.0000000	0.0000000	19.308943	1.615396e-03	-3.153112
PhysicalActivity=PhysicalActivity_No	1.6393443	10.344828	37.195122	1.129509e-03	-3.256108
GenHealth=Fair	0.0000000	0.0000000	23.373984	3.487737e-04	-3.582153
DiffWalking=DiffWalking_Yes	0.6060606	3.448276	33.536585	8.320653e-05	-3.934972
GenHealth=Good	0.0000000	0.0000000	38.414634	4.563273e-07	-5.043821

La classe 1 possède tous les individus se percevant comme ayant une excellente santé .La modalité “Excellent” de la variable “GenHealth” a un faible effectif comparé aux autres modalités (cf univariée) ce qui explique pourquoi cette classe contient le moins d’individus .

Les autres variables caractérisant cette classe semblent cohérents .96% des individus ici n’ont pas une mobilité réduite ,peuvent monter des surfaces en pente.

De même 90% d’entre eux disent ne pas avoir de problèmes physiques , blessures récentes ,prises excessives de médicaments récemment ..., et pratiquent une activité physique en plus de leurs allers/retours au travail.

## Classe 2

	Cla/Mod	Mod/Cla	Global	p.value	v.test
GenHealth=Very good	88.764045	91.860465	18.089431	3.363715e-71	17.841528
PhysicalHealth=PhysicalHealth_Aucun jour	26.848249	80.232558	52.235772	4.630477e-09	5.859935
Smoking=Smoking_No	29.047619	70.930233	42.682927	6.863982e-09	5.794219
PhysicalActivity=PhysicalActivity_Yes	23.948220	86.046512	62.804878	2.521244e-07	5.156116
DiffWalking=DiffWalking_No	22.935780	87.209302	66.463415	2.215190e-06	4.732731
MentalHealth=MentalHealth_Aucun jour	21.987952	84.883721	67.479675	7.696991e-05	3.953644
SkinCancer=SkinCancer_Yes	34.328358	26.744186	13.617886	3.145174e-04	3.603040
Diabetic=Diabetic_No	20.909091	80.232558	67.073171	3.447823e-03	2.924706
AlcoholDrinking=AlcoholDrinking_No	18.454936	100.000000	94.715447	5.853143e-03	2.755897
SleepTime=6h à 8h	20.748299	70.930233	59.756098	1.931072e-02	2.339478
MentalHealth=MentalHealth_Nombreux jours	9.756098	13.953488	25.000000	7.146316e-03	-2.689948
MentalHealth=MentalHealth_Un mois	2.702703	1.162791	7.520325	6.401346e-03	-2.726482
AlcoholDrinking=AlcoholDrinking_Yes	0.000000	0.000000	5.284553	5.853143e-03	-2.755897
PhysicalHealth=PhysicalHealth_Nombreux jours	10.000000	16.279070	28.455285	4.564571e-03	-2.836257
Diabetic=Diabetic_Yes	10.493827	19.767442	32.926829	3.447823e-03	-2.924706
GenHealth=Excellent	0.000000	0.000000	5.894309	3.169180e-03	-2.950833
SleepTime=9h et plus	0.000000	0.000000	6.910569	1.128739e-03	-3.256301
SkinCancer=SkinCancer_No	14.823529	73.255814	86.382114	3.145174e-04	-3.603040
GenHealth=Poor	2.857143	2.325581	14.227642	8.694892e-05	-3.924392
PhysicalHealth=PhysicalHealth_Un mois	3.157895	3.488372	19.308943	4.372875e-06	-4.592824
DiffWalking=DiffWalking_Yes	6.666667	12.790698	33.536585	2.215190e-06	-4.732731
PhysicalActivity=PhysicalActivity_No	6.557377	13.953488	37.195122	2.521244e-07	-5.156116
Smoking=Smoking_Yes	8.865248	29.069767	57.317073	6.863982e-09	-5.794219
GenHealth=Fair	1.739130	2.325581	23.373984	5.203159e-09	-5.840541
GenHealth=Good	1.587302	3.488372	38.414634	2.280309e-16	-8.206341

La classe 2 est principalement composée d'individus prétendant être en très bonne santé ("Very Good") (88%). Donc une grande majorité semble très bien se porter globalement malgré des problèmes liés au cœur.

De même, plus de 80% des individus ici n'ont signalé aucun trouble émotionnel, ou une diminution des capacités physiques qui impacterait fortement leur vie au quotidien, donc un retour aussi positif sur leur santé récente.

Enfin, une part non négligeable des individus qui ont eu un cancer de la peau sont concentrés ici (pratiquement un-tiers).



### Classe 3

	Cla/Mod	Mod/Cla	Global	p.value	v.test
GenHealth=Good	82.539683	65.0000000	38.414634	5.685089e-34	12.150737
AlcoholDrinking=AlcoholDrinking_No	51.502146	100.0000000	94.715447	1.436414e-08	5.668990
GenHealth=Fair	71.304348	34.1666667	23.373984	2.857926e-08	5.549925
PhysicalHealth=PhysicalHealth_Nombreux jours	63.571429	37.0833333	28.455285	3.548699e-05	4.135055
DiffWalking=DiffWalking_No	54.128440	73.7500000	66.463415	8.469617e-04	3.336973
SleepTime=6h à 8h	53.741497	65.8333333	59.756098	7.462509e-03	2.675468
Stroke=Stroke_No	50.704225	90.0000000	86.585366	3.074308e-02	2.160382
Diabetic=Diabetic_Yes	55.555556	37.5000000	32.926829	3.592411e-02	2.097785
Diabetic=Diabetic_No	45.454545	62.5000000	67.073171	3.592411e-02	-2.097785
Stroke=Stroke_Yes	36.363636	10.0000000	13.414634	3.074308e-02	-2.160382
DiffWalking=DiffWalking_Yes	38.181818	26.2500000	33.536585	8.469617e-04	-3.336973
MentalHealth=MentalHealth_Un mois	18.918919	2.9166667	7.520325	1.196688e-04	-3.846804
AlcoholDrinking=AlcoholDrinking_Yes	0.000000	0.0000000	5.284553	1.436414e-08	-5.668990
GenHealth=Excellent	0.000000	0.0000000	5.894309	1.627278e-09	-6.031219
PhysicalHealth=PhysicalHealth_Un mois	20.000000	7.9166667	19.308943	1.688906e-10	-6.387254
SleepTime=9h et plus	0.000000	0.0000000	6.910569	4.119878e-11	-6.599701
GenHealth=Poor	0.000000	0.0000000	14.227642	1.960030e-23	-9.975050
GenHealth=Very good	2.247191	0.8333333	18.089431	1.099812e-26	-10.692815

La classe 3 contient principalement des individus se disant avoir une bonne santé en général("Good") mais une part non négligeable d'entre eux (34%) disent que celle ci est assez moyenne , c'est à dire qu'ils peuvent avoir un peu de mal à accomplir physiquement ou même socialement ce qu'on exige d'une personne "normal".

D'ailleurs Cette classe contient 63% des individus réclamant avoir quelques soucis physiques durant de nombreux jours.

De même , on retrouve ici un peu plus de la moitié des individus avec un trouble de la marche au quotidien .

Enfin , tous les individus ici ne sont pas de grands consommateurs d'alcool soit 100% , contre 51% dans la population totale.

### Classe 4

	Cla/Mod	Mod/Cla	Global	p.value	v.test
AlcoholDrinking=AlcoholDrinking_Yes	92.307692	100.00000	5.284553	8.811086e-39	13.025078
BMI=Surpoids	8.465608	66.66667	38.414634	4.878711e-03	2.814935
Smoking=Smoking_Yes	7.092199	83.33333	57.317073	7.008343e-03	2.696448
MentalHealth=MentalHealth_Un mois	13.513514	20.83333	7.520325	3.295035e-02	2.132688
AgeCategory=moins de 50 ans	12.500000	20.83333	8.130081	4.572217e-02	1.997949
BMI=Obésité/obésité morbide	2.538071	20.83333	40.040650	4.768929e-02	-1.980127
AgeCategory=65 or more	3.144654	41.66667	64.634146	2.085666e-02	-2.310569
Smoking=Smoking_No	1.904762	16.66667	42.682927	7.008343e-03	-2.696448
AlcoholDrinking=AlcoholDrinking_No	0.000000	0.000000	94.715447	8.811086e-39	-13.025078

Cette classe se caractérise par les personnes qui boivent beaucoup d'alcool malgré qu'ils soient atteints de maladies cardiovasculaires. Dans notre analyse univariée nous avons vu que cette modalité reporte un faible effectif de la population étudiés . Cette classe se caractérise donc par un profil atypique d'individus sur cette variable.

## Classe 5

	Cla/Mod	Mod/Cla	Global	p.value	v.test
SleepTime=9h et plus	88.235294	100.00000	6.910569	5.283153e-44	13.912980
AgeCategory=65 or more	8.176101	86.66667	64.634146	6.781022e-03	2.707411
SleepTime=2h à 5h	0.000000	0.00000	33.333333	3.274139e-06	-4.652828
SleepTime=6h à 8h	0.000000	0.00000	59.756098	3.368985e-13	-7.278762

La classe 5 est celle qui regroupe tous les individus qui dorment plus de 9 heures par jour soit 100% de ces individus contre 88% de la population totale.

## Classe 6

	Cla/Mod	Mod/Cla	Global	p.value	v.test
GenHealth=Poor	84.285714	71.084337	14.227642	4.174074e-45	14.093330
PhysicalHealth=PhysicalHealth_Un mois	65.263158	74.698795	19.308943	3.852212e-36	12.552486
DiffWalking=DiffWalking_Yes	41.818182	83.132530	33.536585	1.118478e-24	10.255455
MentalHealth=MentalHealth_Un mois	59.459459	26.506024	7.520325	1.027962e-09	6.105007
PhysicalActivity=PhysicalActivity_No	28.415301	62.650602	37.195122	2.727917e-07	5.141336
Smoking=Smoking_Yes	22.340426	75.903614	57.317073	1.336216e-04	3.819686
SleepTime=2h à 5h	26.219512	51.807229	33.333333	1.446698e-04	3.800045
KidneyDisease=KidneyDisease_Yes	35.185185	22.891566	10.975610	5.247515e-04	3.467793
Stroke=Stroke_Yes	31.818182	25.301205	13.414634	1.258112e-03	3.225367
AlcoholDrinking=AlcoholDrinking_No	17.811159	100.000000	94.715447	7.131484e-03	2.690641
SleepTime=9h et plus	2.941176	1.204819	6.910569	1.392920e-02	-2.459084
SleepTime=6h à 8h	13.265306	46.987952	59.756098	1.035793e-02	-2.563646
AlcoholDrinking=AlcoholDrinking_Yes	0.000000	0.000000	5.284553	7.131484e-03	-2.690641
GenHealth=Excellent	0.000000	0.000000	5.894309	3.953759e-03	-2.881828
Stroke=Stroke_No	14.553991	74.698795	86.585366	1.258112e-03	-3.225367
KidneyDisease=KidneyDisease_No	14.611872	77.108434	89.024390	5.247515e-04	-3.467793
Smoking=Smoking_No	9.523810	24.096386	42.682927	1.336216e-04	-3.819686
MentalHealth=MentalHealth_Aucun jour	11.144578	44.578313	67.479675	2.406243e-06	-4.715917
GenHealth=Very good	1.123596	1.204819	18.089431	2.792850e-07	-5.136915
PhysicalActivity=PhysicalActivity_Yes	10.032362	37.349398	62.804878	2.727917e-07	-5.141336
GenHealth=Good	4.232804	9.638554	38.414634	2.329030e-10	-6.337910
DiffWalking=DiffWalking_No	4.281346	16.867470	66.463415	1.118478e-24	-10.255455
PhysicalHealth=PhysicalHealth_Aucun jour	1.167315	3.614458	52.235772	1.018463e-25	-10.484441

Cette classe est principalement constituée de personnes avec un retour négatif sur leur santé globale( 71% contre 84% de la population totale).

74% des individus disent avoir souffert sur le plan physique récemment ,lésions physiques...,et une part non négligeable (26%) sur le plan émotionnelle (stress,anxiété..).

La classe 6 est aussi principalement constituée des individus qui fument beaucoup (75%) et qui ne pratiquent aucun sport ou autre dépense physique (62%) , de même 83% ont une mobilité réduite .

On constate aussi que la moitié dort entre 2 heures et 5 heures ce qui est peu , peu importe la tranche d'âge .

Enfin , une part non négligeable d'individus ont fait un AVC (25%) au cours de leur vie et ont eu une maladie des reins, soit 22% des sujets ici ont un facteur de risque cardiovasculaire liée à ce type de maladie. Ici on a un regroupement des individus atypiques sur ces deux variables ( cf analyse univariée , faible effectif sur "Stroke\_Yes" , "Kidney\_Yes").

La classe 6 semble avoir des individus avec une très mauvaise santé , avec une hygiène de vie non recommandée pour leur cas.

La classification non supervisée nous a permis de mettre en avant 6 groupes d'individus ou la qualité de vie (santé globale et récente) semble décroître de la classe 1,2,3,6 . De même, certaines classes possèdent des individus atypiques sur certaines variables liées aux antécédents médicaux et au mode de vie . Enfin la classe 3 possède la plupart des individus qui se perçoivent comme ayant une santé normale voir moyenne ,ce qui explique pourquoi elle possède le plus grand effectif ,vu que ce sont les deux modalités de "GenHealth" avec le plus grand nombre.

En annexe (annexe 12 ) on retrouve les individus les plus proches et les plus éloignés du centre de gravité de chaque classe .Les modalités prises par leurs variables semblent cohérentes avec la description des classes et n'appellent à aucune remarque particulière.

## Conclusion

L'analyse de ce jeu de données nous a permis de mettre en avant différents groupes de sujets atteints de maladies cardiovasculaires.

L'analyse bivariée nous a permis de voir quelques relations intéressantes entre le sex ,l'âge, la race des sujets et leurs antécédents médicaux qui auraient pu jouer un rôle déterminant dans leur état actuel qui nous intéresse.

L'analyse des correspondances multiples a permis de résumer ce qui oppose les différents sujets sur la base des différentes variables choisies .

Les variables liées aux “conditions physiques/mentales” ( GenHealth,MentalHealth,PhysicalHealth,DiffWalking) sont les plus significatives dans la disparité des individus.

On a de même pu observer que les hommes âgés de race blanche ont une meilleure santé en dépit de leur situation (maladie cardiovasculaire), du moins ils en ont une perception plus positive , contrairement aux femmes plus jeunes (moins de 50 ans ) et généralement issues d'autres races.

La classification non supervisée a permis d'observer des groupes de malades homogènes aux profils similaires en 6 classes allant de ceux qui ont une perception très positive de leur santé et une bonne hygiène de vie à ceux qui au contraire ont une perception négative .

A l'issue de cette étude, on peut dire qu'aux Etats-Unis, les personnes atteintes de maladies cardiovasculaires ont différentes attitudes vis-à- vis de leur santé , une sommation des facteurs de risques cardiovasculaires différentes notamment pour les antécédents médicaux.

D'autres facteurs pourraient être intéressant pour enrichir cette étude comme le métier , le statut matrimonial , le niveau scolaire , l'historique familiale .

## Références :

Personal key of Heart disease dataset.URL:

<https://www.kaggle.com/datasets/kamilpytlak/personal-key-indicators-of-heart-disease>

Audigier ,V.(2022,février 8). STA101:Analyse des correspondances multiples.URL:

[https://par.moodle.lecnam.net/pluginfile.php/940295/mod\\_label/intro/acm1.pdf](https://par.moodle.lecnam.net/pluginfile.php/940295/mod_label/intro/acm1.pdf)

François Husson, S.I,(2016).Analyse des données avec R.Presse universitaires de Rennes.

CDC's Healthy Days Measures Used in the America's Health Rankings.URL:

<https://www.cdc.gov/hrqol/featured-items/healthy-days.htm>

Méthode ACM avec R.URL:

<http://www.sthda.com/french/articles/38-methodes-des-composantes-principales-dans-r-guide-pratique/84-acm-dans-r-avec-factor-miner-scripts-faciles-et-cours/#gestion-des-donnees-manquantes>

## IV. Annexes

Notons que tout a été généré à partir du code R et des packages associés (Factoshiny Factominer..) (y compris les tableau de fréquences et profils-lignes en analyse univariée et bivariée)

### Partie Analyse Univariée

#### Annexe 1 :

Tableau de fréquences des variables qualitatives :

<b>-stroke</b> n % val% No 432 86.4 86.4 Yes 68 13.6 13.6	<b>-PhysicalActivities</b> n % val% No 187 37.4 37.4 Yes 313 62.6 62.6	<b>-Sex</b> n % val% Female 206 41.2 41.2 Male 294 58.8 58.8	<b>-Smoking</b> n % val% No 212 42.4 42.4 Yes 288 57.6 57.6
<b>-PhysicalHealth</b> n % val% 0 260 52.0 52.0 1 9 1.8 1.8 2 19 3.8 3.8 3 15 3.0 3.0 4 6 1.2 1.2 5 15 3.0 3.0 6 4 0.8 0.8 7 8 1.6 1.6 8 3 0.6 0.6 9 1 0.2 0.2 10 14 2.8 2.8 12 1 0.2 0.2 14 6 1.2 1.2 15 18 3.6 3.6 17 1 0.2 0.2 19 1 0.2 0.2 20 11 2.2 2.2 21 4 0.8 0.8 25 3 0.6 0.6 28 1 0.2 0.2 30 100 20.0 20.0	<b>-MentalHealth</b> n % val% 0 335 67.0 67.0 1 8 1.6 1.6 2 23 4.6 4.6 3 14 2.8 2.8 4 7 1.4 1.4 5 18 3.6 3.6 7 5 1.0 1.0 8 1 0.2 0.2 10 15 3.0 3.0 12 1 0.2 0.2 14 5 1.0 1.0 15 16 3.2 3.2 20 8 1.6 1.6 23 1 0.2 0.2 25 3 0.6 0.6 30 40 8.0 8.0	<b>-AgeCategory</b> n % val% 18-24 1 0.2 0.2 25-29 4 0.8 0.8 30-34 6 1.2 1.2 35-39 10 2.0 2.0 40-44 7 1.4 1.4 45-49 12 2.4 2.4 50-54 25 5.0 5.0 55-59 45 9.0 9.0 60-64 66 13.2 13.2 65-69 81 16.2 16.2 70-74 76 15.2 15.2 75-79 72 14.4 14.4 80 or older 95 19.0 19.0	<b>-SleepTime</b> n % val% 2 3 0.6 0.6 3 4 0.8 0.8 4 23 4.6 4.6 5 41 8.2 8.2 6 96 19.2 19.2 7 110 22.0 22.0 8 162 32.4 32.4 9 27 5.4 5.4 10 21 4.2 4.2 11 1 0.2 0.2 12 7 1.4 1.4 14 1 0.2 0.2 15 2 0.4 0.4 16 1 0.2 0.2 22 1 0.2 0.2
<b>-KidneyDisease</b> n % val% No 446 89.2 89.2 Yes 54 10.8 10.8	<b>-AlcoholDrinking</b> n % val% No 474 94.8 94.8 Yes 26 5.2 5.2	<b>-Race</b> n % val% American Indian 16 3.2 3.2 Asian 4 0.8 0.8 Black 31 6.2 6.2 Hispanic 35 7.0 7.0 Other 17 3.4 3.4 White 397 79.4 79.4	<b>-Genhealth</b> n % val% Excellent 30 6.0 6.0 Fair 117 23.4 23.4 Good 190 38.0 38.0 Poor 73 14.6 14.6 Very good 90 18.0 18.0

<b>-SkinCancer</b> n % val% No 432 86.4 86.4 Yes 68 13.6 13.6	<b>-Asthma</b> n % val% No 402 80.4 80.4 Yes 98 19.6 19.6	<b>-Diabetic</b> n % val% No 323 64.6 64.6 No, borderline diabetes 13 2.6 2.6 Yes 161 32.2 32.2 Yes (during pregnancy) 3 0.6 0.6	
--	--	---	--

## Annexe 2 :

test de normalité pour BMI :

shapiro.test(rand\_heart\$BMI)-----> data: rand\_heart1\$BMI W = 0.92512, p-value = 4.394e-15

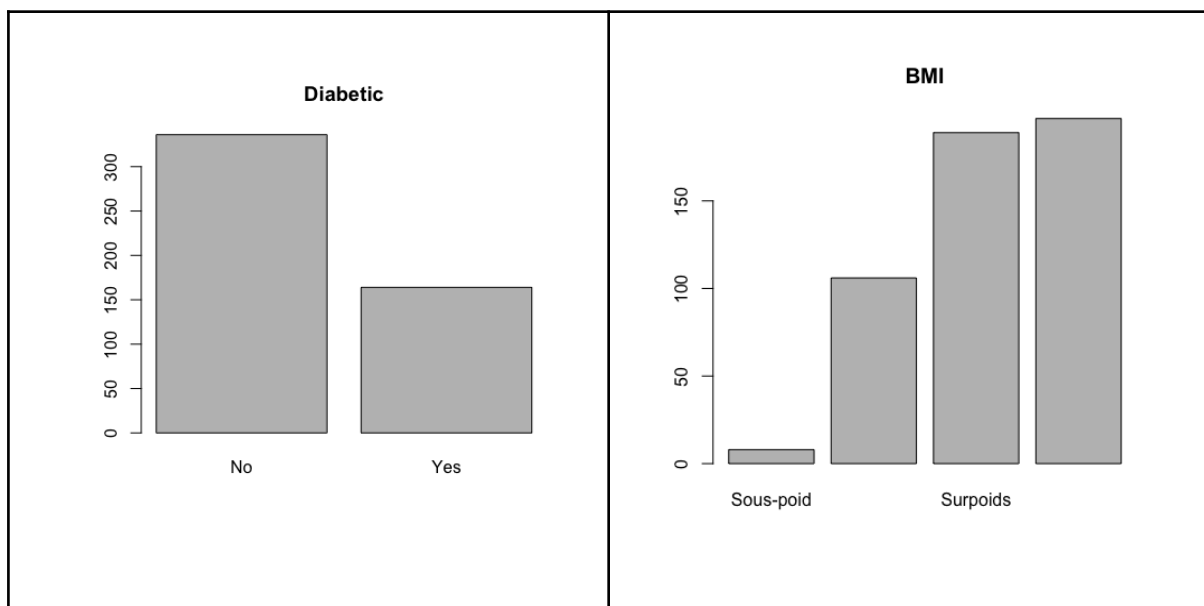
## Annexe 3 :

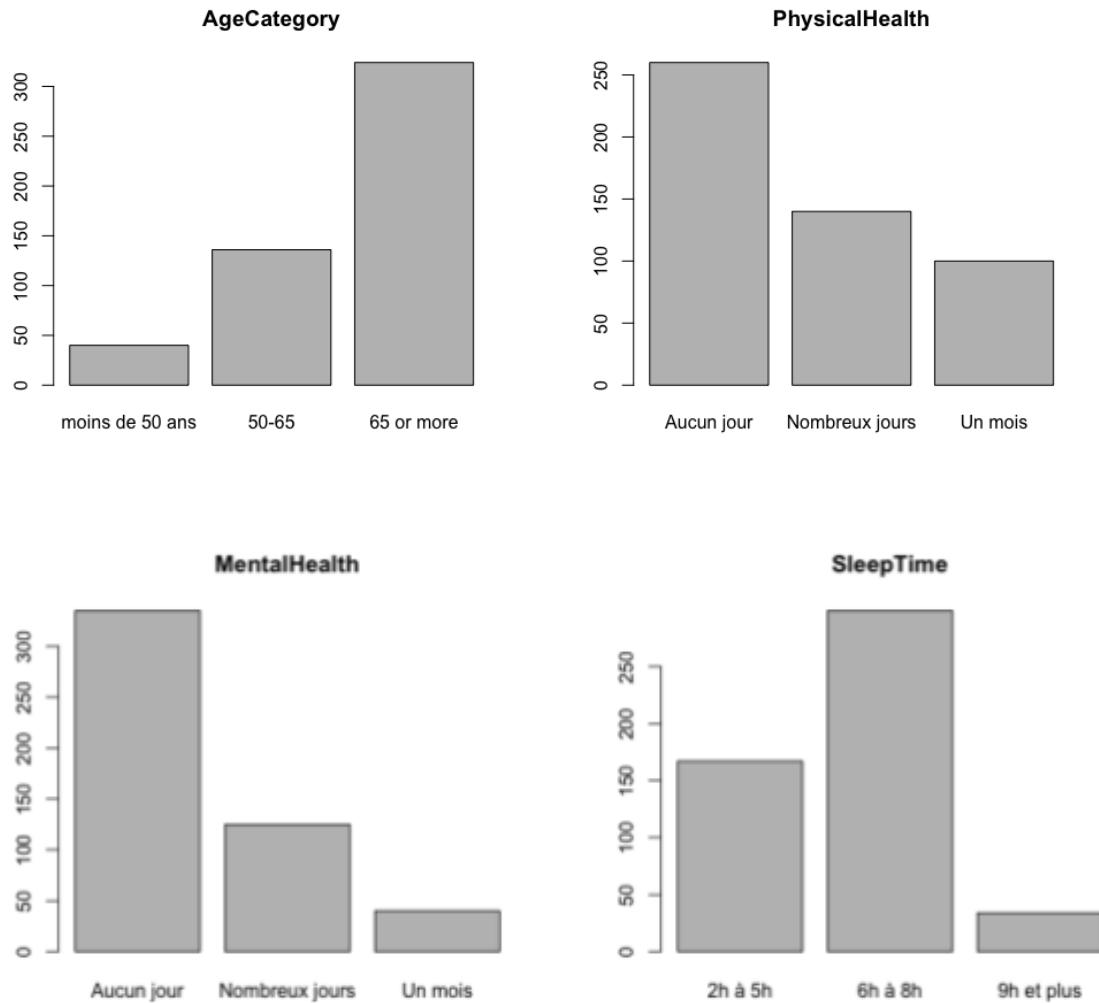
boxplot pour BMI :

## Annexe 4 :

Les diagrammes en barre des variables après leur discrétisation:

(AgeCategory,BMI,SleepTime,MentalHealth,PhysicalHealth,Race,Diabetic)





## Partie Analyse Bivariée

### Annexe 5:

#### Test du khi deux et test exact de fisher

##### **-AgeCategory-Asthmatic:**

X-squared = 8.694, df = 2, p-value = 0.01295

##### **-AgeCategory-SkinCancer:**

p-value = 0.01109

##### **-AgeCategory-BMI:**

p-value = 0.004308

alternative hypothesis: two.sided

##### **-AgeCategory-Diabetic AgeCategory-KidneyDisease AgeCategory-Stroke**

X-squared = 3.2914, df = 2, p-value = 0.1929

p-value = 0.3098



alternative hypothesis: two.sided

X-squared = 5.2186, df = 2, p-value = 0.07359

---

**-Race-Diabetic**

X-squared = 5.2369, df = 1, p-value = 0.02211

**-Race-KidneyDisease**

X-squared = 6.9055, df = 1, p-value = 0.008593

**-Race-BMI Race-SkinCancer Race-Stroke**

p-value = 0.4325      X-squared = 3.157, df = 1, p-value = 0.0756

X-squared = 0.23165, df = 1, p-value = 0.6303

---

**-Sex-Asthma**

X-squared = 4.3613, df = 1, p-value = 0.03676

**-Sex -Diabetic Sex-Stroke Sex-SkinCancer Sex-BMI Sex-KidneyDisease**

p-value = 1

X-squared = 0.86856, df = 1, p-value = 0.3514

X-squared = 0.44476, df = 1, p-value = 0.5048

p-value = 0.3237

X-squared = 2.3638, df = 1, p-value = 0.1242

**Annexe 6:**

Tableaux des profils lignes

<b>- AgeCategory-BMI:</b>					
	Sous-poid	poid-Normal	Surpoids	Obésité/obésité morbide	Total
moins de 50 ans	0.0	25.0	25.0	50.0	100.0
50-65	1.5	14.0	33.1	51.5	100.0
65 or more	1.9	23.8	41.4	33.0	100.0
Ensemble	1.6	21.2	37.8	39.4	100.0

<div>-AgeCategory-Asthmatic:</div> <table><tr><td></td><td>No</td><td>Yes</td><td>Total</td></tr><tr><td>moins de 50 ans</td><td>67.5</td><td>32.5</td><td>100.0</td></tr><tr><td>50-65</td><td>75.7</td><td>24.3</td><td>100.0</td></tr><tr><td>65 or more</td><td>84.0</td><td>16.0</td><td>100.0</td></tr><tr><td>Ensemble</td><td>80.4</td><td>19.6</td><td>100.0</td></tr></table>		No	Yes	Total	moins de 50 ans	67.5	32.5	100.0	50-65	75.7	24.3	100.0	65 or more	84.0	16.0	100.0	Ensemble	80.4	19.6	100.0	<div>-AgeCategory-SkinCancer:</div> <table><tr><td></td><td>No</td><td>Yes</td><td>Total</td></tr><tr><td>moins de 50 ans</td><td>97.5</td><td>2.5</td><td>100.0</td></tr><tr><td>50-65</td><td>90.4</td><td>9.6</td><td>100.0</td></tr><tr><td>65 or more</td><td>83.3</td><td>16.7</td><td>100.0</td></tr><tr><td>Ensemble</td><td>86.4</td><td>13.6</td><td>100.0</td></tr></table>		No	Yes	Total	moins de 50 ans	97.5	2.5	100.0	50-65	90.4	9.6	100.0	65 or more	83.3	16.7	100.0	Ensemble	86.4	13.6	100.0	<div>-Race-Diabetic</div> <table><tr><td></td><td>No</td><td>Yes</td><td>Total</td></tr><tr><td>Other</td><td>57.3</td><td>42.7</td><td>100.0</td></tr><tr><td>White</td><td>69.8</td><td>30.2</td><td>100.0</td></tr><tr><td>Ensemble</td><td>67.2</td><td>32.8</td><td>100.0</td></tr></table>		No	Yes	Total	Other	57.3	42.7	100.0	White	69.8	30.2	100.0	Ensemble	67.2	32.8	100.0
	No	Yes	Total																																																							
moins de 50 ans	67.5	32.5	100.0																																																							
50-65	75.7	24.3	100.0																																																							
65 or more	84.0	16.0	100.0																																																							
Ensemble	80.4	19.6	100.0																																																							
	No	Yes	Total																																																							
moins de 50 ans	97.5	2.5	100.0																																																							
50-65	90.4	9.6	100.0																																																							
65 or more	83.3	16.7	100.0																																																							
Ensemble	86.4	13.6	100.0																																																							
	No	Yes	Total																																																							
Other	57.3	42.7	100.0																																																							
White	69.8	30.2	100.0																																																							
Ensemble	67.2	32.8	100.0																																																							
<div>-Race-Asthma</div> <table><tr><td></td><td>No</td><td>Yes</td><td>Total</td></tr><tr><td>Other</td><td>64.1</td><td>35.9</td><td>100.0</td></tr><tr><td>White</td><td>84.6</td><td>15.4</td><td>100.0</td></tr><tr><td>Ensemble</td><td>80.4</td><td>19.6</td><td>100.0</td></tr></table>		No	Yes	Total	Other	64.1	35.9	100.0	White	84.6	15.4	100.0	Ensemble	80.4	19.6	100.0	<div>-Race-KidneyDisease</div> <table><tr><td></td><td>No</td><td>Yes</td><td>Total</td></tr><tr><td>Other</td><td>81.6</td><td>18.4</td><td>100.0</td></tr><tr><td>White</td><td>91.2</td><td>8.8</td><td>100.0</td></tr><tr><td>Ensemble</td><td>89.2</td><td>10.8</td><td>100.0</td></tr></table>		No	Yes	Total	Other	81.6	18.4	100.0	White	91.2	8.8	100.0	Ensemble	89.2	10.8	100.0	<div>-Sex-Asthma</div> <table><tr><td></td><td>No</td><td>Yes</td><td>Total</td></tr><tr><td>Female</td><td>75.7</td><td>24.3</td><td>100.0</td></tr><tr><td>Male</td><td>83.7</td><td>16.3</td><td>100.0</td></tr><tr><td>Ensemble</td><td>80.4</td><td>19.6</td><td>100.0</td></tr></table>		No	Yes	Total	Female	75.7	24.3	100.0	Male	83.7	16.3	100.0	Ensemble	80.4	19.6	100.0								
	No	Yes	Total																																																							
Other	64.1	35.9	100.0																																																							
White	84.6	15.4	100.0																																																							
Ensemble	80.4	19.6	100.0																																																							
	No	Yes	Total																																																							
Other	81.6	18.4	100.0																																																							
White	91.2	8.8	100.0																																																							
Ensemble	89.2	10.8	100.0																																																							
	No	Yes	Total																																																							
Female	75.7	24.3	100.0																																																							
Male	83.7	16.3	100.0																																																							
Ensemble	80.4	19.6	100.0																																																							

## Partie ACM

## Annexe 7:

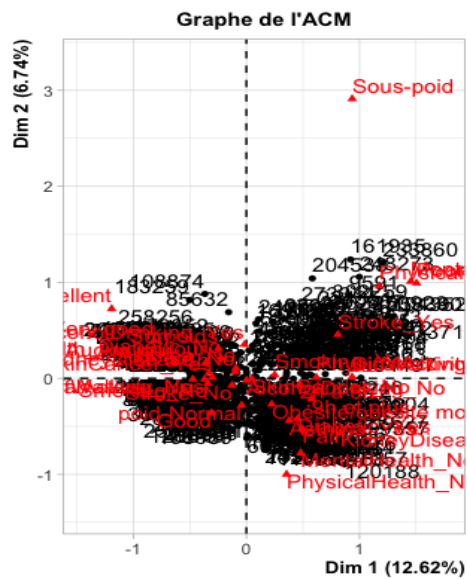


Fig-1 : Nuage des individus avec la modalité “Sous-Poid”

## Annexe 8:

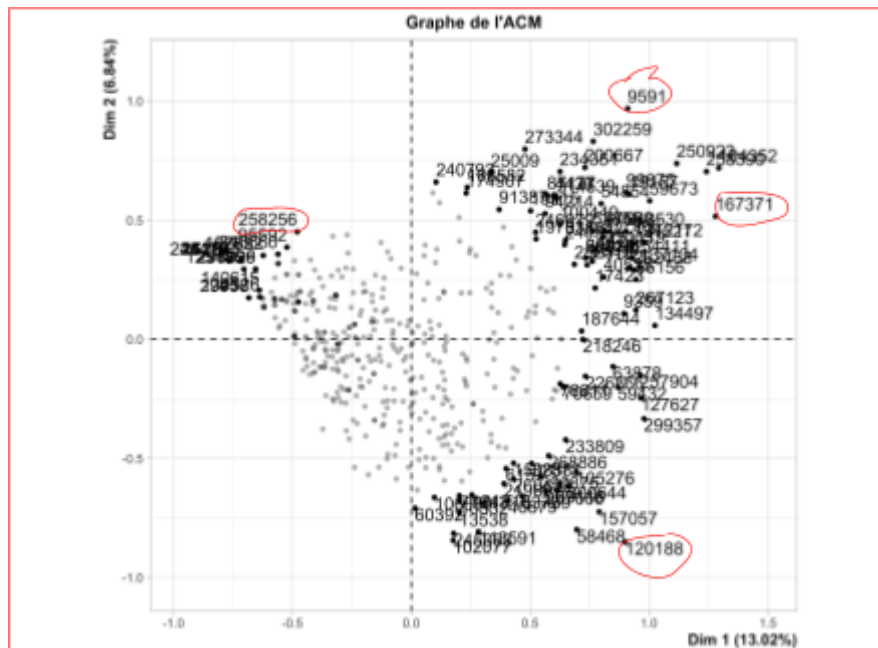


Fig-2 : Nuage des individus qui contribuent axe 1:2

## Annexe 9 :



Fig-3 : Nuage des individus qui contribuent axe 3:4

## Annexe 10

	eigenvalue	percentage of variance	cumulative percentage of variance
dim 1	0.19533571	13.022381	13.02238
dim 2	0.10256779	6.837853	19.86023
dim 3	0.09697111	6.464741	26.32497
dim 4	0.09021421	6.014281	32.33925
dim 5	0.08442678	5.628452	37.96771
dim 6	0.08044786	5.363191	43.33090
dim 7	0.07383493	4.922329	48.25323
dim 8	0.07064181	4.709454	52.96268
dim 9	0.07016096	4.677397	57.64008
dim 10	0.06871786	4.581191	62.22127
dim 11	0.06733811	4.489207	66.71048
dim 12	0.06543743	4.362496	71.07297
dim 13	0.06017713	4.011809	75.08478
dim 14	0.05973287	3.982191	79.06697
dim 15	0.05729539	3.819693	82.88666
dim 16	0.05680067	3.786711	86.67338
dim 17	0.05055365	3.370244	90.04362
dim 18	0.04450240	2.966826	93.01045
dim 19	0.04248541	2.832361	95.84281
dim 20	0.03620785	2.413857	98.25666
dim 21	0.02615005	1.743336	100.00000

Tab.2-Les pourcentages d'inerties et les valeurs propres des dimensions

## Annexe 11 :

### Descriptions automatiques des axes : Tests de Fisher et tests de student

	R2	p.value
GenHealth	6.183E-01	2.087E-100
PhysicalHealth	5.118E-01	7.144E-77
DiffWalking	4.553E-01	1.226E-66
MentalHealth	2.888E-01	6.449E-37
PhysicalActivity	2.222E-01	1.382E-28
SleepTime	1.687E-01	2.408E-20
Diabetic	1.010E-01	5.278E-13
KidneyDisease	9.520E-02	2.600E-12
Stroke	9.121E-02	7.789E-12
Smoking	8.512E-02	4.136E-11
BMI	4.698E-02	7.774E-06
Asthma	3.910E-02	9.946E-06
Sex	3.426E-02	3.614E-05
Race	3.009E-02	1.102E-04
AgeCategory	1.613E-02	1.875E-02
SkinCancer	1.095E-02	2.028E-02

	Estimate	p.value
DiffWalking=DiffWalking_Yes	3.158E-01	1.226E-66
GenHealth=Poor	6.892E-01	2.120E-51
PhysicalHealth=PhysicalHealth_Un mois	3.777E-01	6.473E-43
PhysicalActivity=PhysicalActivity_No	2.155E-01	1.382E-28
MentalHealth=MentalHealth_Un mois	4.012E-01	3.216E-21
SleepTime=2h à 5h	9.960E-02	1.781E-14
Diabetic=Diabetic_Yes	1.494E-01	5.278E-13
KidneyDisease=KidneyDisease_Yes	2.181E-01	2.600E-12
Stroke=Stroke_Yes	1.958E-01	7.789E-12
Smoking=Smoking_Yes	1.304E-01	4.136E-11
GenHealth=Fair	2.518E-01	2.566E-10
PhysicalHealth=PhysicalHealth_Nombreux jours	3.681E-02	5.133E-08
BMI=Obésité/obésité morbide	1.123E-01	6.655E-06
Asthma=Asthma_Yes	1.103E-01	9.946E-06
Sex=Female	8.315E-02	3.614E-05
Race=Other	9.525E-02	1.102E-04
SleepTime=9h et plus	1.595E-01	2.068E-04
SkinCancer=SkinCancer_No	6.741E-02	2.028E-02
AgeCategory=moins de 50 ans	1.004E-01	2.284E-02
SkinCancer=SkinCancer_Yes	-6.741E-02	2.028E-02
AgeCategory=65 or more	-8.798E-02	1.481E-02
Race=White	-9.525E-02	1.102E-04

Sex=Male	-8.315E-02	3.614E-05
BMI=Surpoids	-1.036E-01	1.990E-05
Asthma=Asthma_No	-1.103E-01	9.946E-06
GenHealth=Good	-1.020E-01	2.021E-07
MentalHealth=MentalHealth_Nombreux jours	-1.996E-02	4.021E-10
Smoking=Smoking_No	-1.304E-01	4.136E-11
Stroke=Stroke_No	-1.958E-01	7.789E-12
GenHealth=Excellent	-5.059E-01	7.620E-12
KidneyDisease=KidneyDisease_No	-2.181E-01	2.600E-12
Diabetic=Diabetic_No	-1.494E-01	5.278E-13
GenHealth=Very good	-3.330E-01	1.097E-18
SleepTime=6h à 8h	-2.591E-01	2.608E-21
PhysicalActivity=PhysicalActivity_Yes	-2.155E-01	1.382E-28
MentalHealth=MentalHealth_Aucun jour	-3.812E-01	1.511E-30
PhysicalHealth=PhysicalHealth_Aucun jour	-4.145E-01	2.466E-64
DiffWalking=DiffWalking_No	-3.158E-01	1.226E-66

Tab.3 - Description automatique de l'axe 1

	R2	p.value		Estimate	p.value
PhysicalHealth	5.216E-01	5.216E-79	PhysicalHealth=PhysicalHealth_Un mois	3.091E-01	1.281E-32
GenHealth	3.286E-01	6.042E-41	GenHealth=Poor	2.752E-01	5.457E-25
MentalHealth	2.886E-01	7.007E-37	MentalHealth=MentalHealth_Aucun jour	4.683E-02	1.159E-13
BMI	8.787E-02	1.714E-10	BMI=Surpoids	1.326E-01	4.209E-11
Asthma	7.187E-02	1.519E-09	Asthma=Asthma_No	1.083E-01	1.519E-09
Diabetic	7.151E-02	1.676E-09	Diabetic=Diabetic_No	9.112E-02	1.676E-09
Stroke	2.309E-02	7.214E-04	MentalHealth=MentalHealth_Un mois	2.608E-01	9.477E-09
Race	2.239E-02	8.714E-04	PhysicalHealth=PhysicalHealth_Aucun jour	3.953E-02	2.060E-05
Sex	1.481E-02	6.876E-03	GenHealth=Very good	5.317E-02	3.790E-05
KidneyDisease	1.384E-02	9.003E-03	Stroke=Stroke_Yes	7.139E-02	7.214E-04
AlcoholDrinking	1.315E-02	1.092E-02	Race=White	5.954E-02	8.714E-04
			GenHealth=Excellent	1.021E-01	2.399E-03
			Sex=Male	3.962E-02	6.876E-03
			KidneyDisease=KidneyDisease_No	6.027E-02	9.003E-03
			AlcoholDrinking=AlcoholDrinking_Yes	8.207E-02	1.092E-02
			AlcoholDrinking=AlcoholDrinking_No	-8.207E-02	1.092E-02
			KidneyDisease=KidneyDisease_Yes	-6.027E-02	9.003E-03
			Sex=Female	-3.962E-02	6.876E-03
			BMI=Obésité/obésité morbide	-4.360E-02	1.014E-03
			Race=Other	-5.954E-02	8.714E-04
			Stroke=Stroke_No	-7.139E-02	7.214E-04
			BMI=poid-Normal	-8.896E-02	1.662E-04
			GenHealth=Good	-1.783E-01	4.496E-09
			Diabetic=Diabetic_Yes	-9.112E-02	1.676E-09
			Asthma=Asthma_Yes	-1.083E-01	1.519E-09
			GenHealth=Fair	-2.522E-01	2.143E-12
			MentalHealth=MentalHealth_Nombreux jours	-3.077E-01	1.108E-33
			PhysicalHealth=PhysicalHealth_Nombreux jours	-3.486E-01	2.982E-60

Tab.4- Description automatique de l'axe 2

	R2	p.value		Estimate	p.value
Stroke	2.671E-01	5.904E-35	Stroke=Stroke_Yes	2.361E-01	5.904E-35
BMI	2.358E-01	2.804E-29	Diabetic=Diabetic_No	1.446E-01	2.792E-24
Diabetic	1.904E-01	2.792E-24	BMI=poid-Normal	1.966E-01	1.544E-19
MentalHealth	1.562E-01	9.118E-19	MentalHealth=MentalHealth_Nombreux jours	2.121E-01	6.160E-19
KidneyDisease	1.253E-01	5.786E-16	KidneyDisease=KidneyDisease_No	1.763E-01	5.786E-16
GenHealth	1.067E-01	3.166E-11	PhysicalHealth=PhysicalHealth_Nombreux jours	1.226E-01	2.258E-10
Smoking	7.381E-02	8.979E-10	Smoking=Smoking_Yes	8.552E-02	8.979E-10
PhysicalHealth	8.160E-02	9.148E-10	PhysicalActivity=PhysicalActivity_Yes	8.437E-02	3.687E-09
PhysicalActivity	6.859E-02	3.687E-09	GenHealth=Poor	1.062E-01	4.221E-05
SkinCancer	2.234E-02	8.821E-04	GenHealth=Excellent	1.574E-01	6.111E-04
SleepTime	2.194E-02	4.408E-03	SkinCancer=SkinCancer_No	6.785E-02	8.821E-04
AgeCategory	2.020E-02	6.808E-03	SleepTime=9h et plus	1.142E-01	1.513E-03
			AgeCategory=moins de 50 ans	1.086E-01	1.904E-03
			BMI=Surpoids	2.355E-03	2.922E-02
			MentalHealth=MentalHealth_Un mois	-1.568E-01	1.063E-03
			SkinCancer=SkinCancer_Yes	-6.785E-02	8.821E-04
			PhysicalHealth=PhysicalHealth_Aucun jour	-8.287E-02	4.974E-07
			GenHealth=Fair	-1.765E-01	1.373E-08
			PhysicalActivity=PhysicalActivity_No	-8.437E-02	3.687E-09
			Smoking=Smoking_No	-8.552E-02	8.979E-10
			MentalHealth=MentalHealth_Aucun jour	-5.535E-02	6.330E-10
			KidneyDisease=KidneyDisease_Yes	-1.763E-01	5.786E-16
			BMI=Obésité/obésité morbide	-1.990E-01	3.479E-23
			Diabetic=Diabetic_Yes	-1.446E-01	2.792E-24
			Stroke=Stroke_No	-2.361E-01	5.904E-35

Tab.5- Description automatique de l'axe 3



	R2	p.value		Estimate	p.value
GenHealth	0.430	0.000	Smoking=Smoking_No	0.159	0.000
Smoking	0.274	0.000	SkinCancer=SkinCancer_Yes	0.223	0.000
SkinCancer	0.260	0.000	GenHealth=Very good	0.217	0.000
AlcoholDrinking	0.111	0.000	AlcoholDrinking=AlcoholDrinking_No	0.224	0.000
PhysicalActivity	0.033	0.000	PhysicalActivity=PhysicalActivity_Yes	0.057	0.000
PhysicalHealth	0.040	0.000	PhysicalHealth=PhysicalHealth_Nombreux jours	0.058	0.000
MentalHealth	0.032	0.000	MentalHealth=MentalHealth_Nombreux jours	0.044	0.001
Stroke	0.018	0.003	Stroke=Stroke_Yes	0.059	0.003
Diabetic	0.014	0.008	GenHealth=Poor	0.027	0.003
BMI	0.018	0.011	BMI=poid-Normal	0.064	0.003
KidneyDisease	0.011	0.020	Diabetic=Diabetic_No	0.038	0.008
Sex	0.011	0.022	GenHealth=Excellent	0.066	0.012
Asthma	0.009	0.031	KidneyDisease=KidneyDisease_Yes	0.050	0.020
			Sex=Female	0.031	0.022
			Asthma=Asthma_Yes	0.037	0.031
			Asthma=Asthma_No	-0.037	0.031
			Sex=Male	-0.031	0.022
			KidneyDisease=KidneyDisease_No	-0.050	0.020
			GenHealth=Fair	-0.008	0.011
			Diabetic=Diabetic_Yes	-0.038	0.008
			Stroke=Stroke_No	-0.059	0.003
			MentalHealth=MentalHealth_Aucun jour	-0.073	0.000
			PhysicalActivity=PhysicalActivity_No	-0.057	0.000
			PhysicalHealth=PhysicalHealth_Aucun jour	-0.075	0.000
			AlcoholDrinking=AlcoholDrinking_Yes	-0.224	0.000
			SkinCancer=SkinCancer_No	-0.223	0.000
			Smoking=Smoking_Yes	-0.159	0.000
			GenHealth=Good	-0.301	0.000

Tab.6- Description automatique de l'axe 4

## Résultats sur les modalités : coordonnees contributions cosinus carres

Fair	0.51	-0.56	-0.46	0.21
Good	-0.29	-0.33	0.09	-0.77
Poor	1.50	1.09	0.45	0.33
Very good	-0.82	0.39	-0.15	0.96
2h à 5h	0.48	-0.06	0.02	0.03
6h à 8h	-0.34	0.04	-0.07	0.01
9h et plus	0.61	-0.08	0.52	-0.29
Asthma_No	-0.10	0.13	0.02	-0.05
Asthma_Yes	0.40	-0.54	-0.08	0.20
KidneyDisease_No	-0.11	0.04	0.12	-0.04
KidneyDisease_Yes	0.88	-0.34	-1.01	0.30
SkinCancer_No	0.04	0.00	0.06	-0.20
SkinCancer_Yes	-0.26	-0.00	-0.38	1.29

	Dim 1	Dim 2	Dim 3	Dim 4
poid-Normal	-0.03	-0.32	0.75	0.25
Surpoids	-0.24	0.37	0.12	-0.04
Obésité/obésité morbide	0.25	-0.18	-0.52	-0.10
Smoking_No	-0.34	-0.06	-0.31	0.61
Smoking_Yes	0.25	0.04	0.23	-0.45
AlcoholDrinking_No	-0.00	-0.03	-0.02	0.08
AlcoholDrinking_Yes	0.00	0.49	0.30	-1.41
Stroke_No	-0.12	-0.06	-0.20	-0.05
Stroke_Yes	0.77	0.39	1.31	0.34
PhysicalHealth_Aucun jour	-0.64	0.18	-0.21	-0.18
PhysicalHealth_Nombreux jours	0.38	-1.03	0.45	0.26
PhysicalHealth_Un mois	1.16	1.02	-0.08	0.12
MentalHealth_Aucun jour	-0.34	0.23	-0.19	-0.12
MentalHealth_Nombreux jours	0.48	-0.88	0.67	0.27
MentalHealth_Un mois	1.43	0.89	-0.52	0.22
DiffWalking_No	-0.48	-0.06	0.03	-0.05
DiffWalking_Yes	0.95	0.12	-0.05	0.10
Diabetic_No	-0.22	0.19	0.31	0.08
Diabetic_Yes	0.45	-0.38	-0.62	-0.17
PhysicalActivity_No	0.61	0.08	-0.34	-0.24
PhysicalActivity_Yes	-0.36	-0.05	0.20	0.14
Excellent	-1.21	0.55	0.62	0.45

coordonnées



	Dim 1	Dim 2	Dim 3	Dim 4
poid-Normal	0.01	1.56	8.89	1.08
Surpoids	0.82	3.65	0.44	0.04
Obésité/obésité morbide	0.89	0.91	8.04	0.33
Smoking_No	1.78	0.11	3.12	12.45
Smoking_Yes	1.33	0.08	2.32	9.27
AlcoholDrinking_No	0.00	0.05	0.02	0.47
AlcoholDrinking_Yes	0.00	0.87	0.35	8.35
Stroke_No	0.45	0.22	2.64	0.19
Stroke_Yes	2.89	1.39	17.04	1.21
PhysicalHealth_Aucun jour	7.74	1.21	1.77	1.41
PhysicalHealth_Nombreux jours	1.54	21.01	4.16	1.50
PhysicalHealth_Un mois	9.43	14.10	0.08	0.22
MentalHealth_Aucun jour	2.81	2.41	1.80	0.81
MentalHealth_Nombreux jours	2.11	13.49	8.24	1.41
MentalHealth_Un mois	5.65	4.19	1.47	0.28
DiffWalking_No	5.58	0.16	0.03	0.14
DiffWalking_Yes	11.07	0.31	0.06	0.28
Diabetic_No	1.22	1.64	4.62	0.37
Diabetic_Yes	2.48	3.34	9.40	0.76
PhysicalActivity_No	5.10	0.17	3.17	1.65
PhysicalActivity_Yes	3.02	0.10	1.88	0.98
Excellent	3.14	1.22	1.64	0.96
Fair	2.20	5.12	3.60	0.80
Good	1.21	2.91	0.21	17.95
Poor	11.66	11.69	2.13	1.20
Very good	4.41	1.94	0.28	13.15
2h à 5h	2.76	0.08	0.01	0.03
6h à 8h	2.47	0.07	0.21	0.01
9h et plus	0.94	0.03	1.40	0.46
Asthma_No	0.28	0.98	0.02	0.15
Asthma_Yes	1.15	4.03	0.09	0.60
KidneyDisease_No	0.38	0.11	1.01	0.10
KidneyDisease_Yes	3.10	0.86	8.21	0.78
SkinCancer_No	0.05	0.00	0.22	2.81
SkinCancer_Yes	0.35	0.00	1.42	17.81

## Contributions

	Dim 1	Dim 2	Dim 3	Dim 4
poid-Normal	0.00	0.03	0.15	0.02
Surpoids	0.04	0.09	0.01	0.00
Obésité/obésité morbide	0.04	0.02	0.18	0.01
Smoking_No	0.09	0.00	0.07	0.27
Smoking_Yes	0.09	0.00	0.07	0.27
AlcoholDrinking_No	0.00	0.01	0.01	0.11
AlcoholDrinking_Yes	0.00	0.01	0.01	0.11
Stroke_No	0.09	0.02	0.27	0.02
Stroke_Yes	0.09	0.02	0.27	0.02
PhysicalHealth_Aucun jour	0.44	0.04	0.05	0.04
PhysicalHealth_Nombreux jours	0.06	0.42	0.08	0.03
PhysicalHealth_Un mois	0.32	0.25	0.00	0.00
MentalHealth_Aucun jour	0.24	0.11	0.08	0.03
MentalHealth_Nombreux jours	0.08	0.26	0.15	0.02
MentalHealth_Un mois	0.17	0.07	0.02	0.00
DiffWalking_No	0.46	0.01	0.00	0.01
DiffWalking_Yes	0.46	0.01	0.00	0.01
Diabetic_No	0.10	0.07	0.19	0.01
Diabetic_Yes	0.10	0.07	0.19	0.01
PhysicalActivity_No	0.22	0.00	0.07	0.03
PhysicalActivity_Yes	0.22	0.00	0.07	0.03
Excellent	0.09	0.02	0.02	0.01
Fair	0.08	0.10	0.06	0.01
Good	0.05	0.07	0.00	0.37
Poor	0.37	0.20	0.03	0.02
Very good	0.15	0.03	0.00	0.20
2h à 5h	0.11	0.00	0.00	0.00
6h à 8h	0.17	0.00	0.01	0.00
9h et plus	0.03	0.00	0.02	0.01
Asthma_No	0.04	0.07	0.00	0.01
Asthma_Yes	0.04	0.07	0.00	0.01
KidneyDisease_No	0.10	0.01	0.13	0.01
KidneyDisease_Yes	0.10	0.01	0.13	0.01
SkinCancer_No	0.01	0.00	0.02	0.26
SkinCancer_Yes	0.01	0.00	0.02	0.26

Cosinus carrés

Tab.8-Résultats sur les variables supplémentaires: : coordonnées, cosinus carrés ,V.test

Dim 1	Dim 2	Dim 3	Dim 4
0.22	-0.15	-0.01	0.12
-0.15	0.10	0.01	-0.09
0.34	-0.21	0.47	-0.16
0.09	-0.06	-0.08	-0.03
-0.08	0.05	-0.02	0.03
0.34	-0.30	-0.16	0.04
-0.09	0.08	0.04	-0.01

Coordonnées

	Dim 1	Dim 2	Dim 3	Dim 4
Female	0.03	0.01	0.00	0.01
Male	0.03	0.01	0.00	0.01
moins de 50 ans	0.01	0.00	0.02	0.00
50-65	0.00	0.00	0.00	0.00
65 or more	0.01	0.00	0.00	0.00
Other	0.03	0.02	0.01	0.00
White	0.03	0.02	0.01	0.00

Cosinus carres

	Dim 1	Dim 2	Dim 3	Dim 4
Female	4.10	-2.70	-0.24	2.29
Male	-4.10	2.70	0.24	-2.29
moins de 50 ans	2.27	-1.36	3.09	-1.04
50-65	1.22	-0.83	-1.14	-0.40
65 or more	-2.43	1.55	-0.71	0.97
Other	3.84	-3.32	-1.75	0.48
White	-3.84	3.32	1.75	-0.48

V.test

## Annexe 12:

Les individus les plus proches du centre de la classe 1,2,3,4,5,6 (paragons)

### Cluster: 1

172273	262343	111535	29916	9456
0.4721585	0.4721585	0.4721585	0.5026275	0.5026275

### Cluster: 2

175126	296795	291639	28848	264285
0.4968995	0.4968995	0.4968995	0.5180747	0.5180747

### Cluster: 3

125123	110383	302921	146451	238072
0.5771343	0.5837284	0.5921292	0.6182770	0.6577083

### Cluster: 4

85358	210751	131242	82774	140695
0.6884204	0.6884204	0.7424176	0.7505206	0.7789227

### Cluster: 5

141147	160385	75809	193005	23712
0.7452373	0.8155284	0.8160438	0.8197087	0.8969179

### Cluster: 6

107890	87177	47560	100449	232500
0.7434835	0.7440856	0.7493693	0.7860358	0.7887675

Les individus les plus éloignés du centre de la classe 1,2,3,4,5,6 (specifiques)

### Cluster: 1

28691	163022	141669	165066	82498
1.687571	1.611097	1.595445	1.581095	1.535006

### Cluster: 2

26622	13538	179352	242642	47085
1.428951	1.412780	1.384270	1.364772	1.363456

### Cluster: 3

112160	219897	10532	83789	146067
1.686061	1.623894	1.575264	1.562398	1.545431

### Cluster: 4

9359	40525	137104	314539	260139
1.913826	1.793695	1.788022	1.750019	1.721665

### Cluster: 5

167552	234351	105276	59650	101836
1.709223	1.700387	1.656476	1.607621	1.589219

### Cluster: 6

104352	250923	248530	258390	302259
1.961872	1.920636	1.785405	1.763622	1.732393

## Annexe 13 :

