MASTER EN SCIENCES MATHÉMATIQUES

Statistiques et problèmes de santé publique

DUMONT, Morgane

Award date:
2014

UNIVERSITE DE NAMUR

Faculté des Sciences

Statistiques et problèmes de santé publique

Promoteur : M. Rémon

Co-Promoteur : Mme Tellier

Mémoire présenté pour l'obtention

du grade académique de master en « sciences mathématiques à finalité spécialisée »

Morgane DUMONT

Juin 2014

Résumé

La santé mentale fait partie des préoccupations des services de santé publique. Qui sont ces personnes qui se rendent pour la première fois dans un service de santé mentale ? Pourquoi font-elles cette démarche et quelles sont leurs attentes ? Y a-t-il des groupes de nouveaux patients avec des caractéristiques semblables ? Nous tentons de répondre à ces questions dans ce mémoire.

L'Observatoire Wallon de la Santé (OWS) nous a fourni des bases de données concernant les nouveaux patients des services de santé mentale wallons arrivés entre 2008 et 2011. Après avoir fixé le contexte de ce mémoire et nettoyé les données, nous analysons le profil, ainsi que l'itinéraire thérapeutique et le problème du patient. Des dépendances sont mises en avant et une classification permet d'identifier différents types de patients.

L'analyse de l'état de santé de la population est très importante pour les services de santé publique afin de communiquer les résultats et d'intervenir dans le secteur pertinent à un moment opportun.

Mots clés : Santé publique, santé mentale, statistiques, dépendances, classifications

Abstract

The public health have to take decisions in order to improve the global health of a population. The health include several fields, such as the mental health. This is the part that we analyse in this paper. Indeed, we use a dataset concerning the new patients of the mental health services in Wallonia to describe the part of the population that decide to consult.

Who are the patient of these services ? Where do they come from ? What are their expectations ? Why do they feel the need to take this step ? A lot of information are available about these new patients and the reasons of this first coming. In this thesis, in order to set the contex of the paper, we first define public health, mental health,... Then, we explain how we have manipulated the first file to obtain an effective data base. We continue with an analysis of the missing values to assure the representativity of our sample.

After these chapters, we focus on the analysis of the data. The first part describes the person's caracteristics, in term of socio-economic and then medical variables. The second part concerns the links between some variables. Finally, we procede to a classification to form groups of people with similar profils.

Keywords : Public health, mental health, statistics, dependences, classifications

Remerciements

Ce mémoire n'aurait pas été possible sans l'intervention et le soutien de certaines personnes. Il est donc important pour moi de formuler ici des remerciements.

Tout d'abord, merci à M. Rémon et Mme Tellier d'avoir mis ce sujet de mémoire sur pied et de m'avoir permis de relever le challenge que représentait un tel sujet. Je tiens à les remercier pour leur disponibilité et leur soutien dans les moments charnières de ce projet.

Il est également important de mentionner l'équipe, créée au sein de l'OWS, me permettant de rencontrer des personnes connaissant les attentes d'une telle étude et expertes dans le domaine de la santé publique et de la santé mentale. Cette équipe était composée de : Hugues Reyniers (Direction des soins ambulatoires DGO5-Service public de Wallonie) ; Christiane Bontemps et Marie Lambert (CRESAM) ; Dominique Dubourg, Anouck Billiet et Véronique Tellier (OWS). Un tout grand merci à chacun d'eux pour leur présence, leurs suggestions, leur motivation, leur soutien et leur enthousiasme.

Du côté mathématique, je souhaite remercier M. Rémon et Johan Barthélemy, qui ont répondu à mes questions d'un point de vue théorique sur les statistiques. Je tiens également à mentionner qu'ils ont tous deux suggérer des solutions intéressantes lorsqu'un problème apparaissait.

Toute ma reconnaissance va également à Simon Pirenne et Annick Seel, qui ont eu la gentillesse de prendre le temps de me relire afin de corriger un maximum d'erreurs orthographiques.

Un grand merci à M. Rémon pour sa relecture suivie de corrections orthographiques et de suggestions permettant d'améliorer le contenu du mémoire.

Table des matières

Introduction

Avec la collaboration de l'Observatoire Wallon de la Santé (OWS), nous allons faire une étude statistique sur des bases de données déjà récoltées concernant la santé mentale en Wallonie. Ces données reprennent des informations sur les premières consultations des patients s'étant rendus dans un service de santé mentale pour la première fois entre 2008 et 2011. Nous analyserons dans ce document uniquement les informations concernant les personnes majeures lors de leur première visite.

Le but de ce mémoire est de répondre à la demande de l'Observatoire Wallon de la Santé. Ils ont demandé une analyse statistique des données collectées par les services de santé mentale subventionnés par la région wallonne. En effet, chaque service de santé mentale a rempli un formulaire pour chaque nouveau patient. Ce formulaire se trouve en annexe. Remarquons que ce questionnaire a été systématiquement complété par le praticien et non par le patient. Ce n'est donc pas une enquête menée au niveau des personnes concernées, mais l'équivalent d'un résumé de chaque nouveau dossier, fait de manière anonyme.

Grâce à ce questionnaire, nous avons des renseignements généraux sur chaque nouveau patient, comme l'année de naissance, le sexe, l'état civil, la nationalité et la langue maternelle, ainsi que des données plus spécifiques telles que le niveau d'étude, le lieu de vie, etc. Enfin, nous avons des données concernant le suivi psychologique de la personne avant et après cette première consultation, ainsi que le type de demande formulée. Chaque service de santé mentale a alors rendu à la Direction Générale Opérationnelle des Pouvoirs Locaux (notée *DGO5*) un fichier comprenant une ligne par nouveau patient. Chaque colonne correspond à la réponse à une question.

Pour réaliser cette étude, nous avons le soutien d'un groupe de travail composé de personnes de divers services : Hugues Reyniers (Direction des soins ambulatoires DGO5-Service public de Wallonie) ; Christiane Bontemps et Marie lambert (CRESAM) ; Dominique Dubourg, Anouck Billiet et Véronique Tellier (OWS). De plus, la jonction des différentes bases de données a déjà été commencée par une ancienne stagiaire de l'OWS : Wyvinne Gaziaux. Ce groupe et moi nous sommes réunis à plusieurs reprises pour diverses raisons. Tout d'abord, ils définissaient les objectifs et les directions des recherches. Ensuite, je leur présentais l'état d'avancement de mes recherches. Dans ce cas, soit l'analyse leur convenait, soit ils la redirigeaient et ajoutaient des exigences. Dans ce mémoire, ce groupe est le commanditaire que nous tentons de satisfaire.

Notons que le gros travail de recherche de ce mémoire n'a pas été de faire des calculs poussés et demandant beaucoup de techniques. Son but était de vulgariser un maximum les analyses faites et de trouver des visualisations et graphiques qui parlaient aux membres du groupe. La demande était de décrire au mieux les données et nous avons dû chercher à représenter un maximum d'informations différentes. Par exemple, lorsqu'une même question permettait plusieurs réponses, il a fallut les décrire sous tous les angles. De plus, pour visualiser les dépendances, nous avons fait 6 sortes différentes de représentations afin de voir laquelle serait la plus pertinente pour l'OWS. Pour appuyer malgré tout les différentes analyses, nous avons ajouté des tests plus statistique.

Étant donné que ce mémoire va se situer dans le contexte large de la santé publique, afin qu'ils soient compréhensibles également pour les mathématiciens, nous allons, dans un premier temps, procéder à un chapitre permettant de bien définir ce milieu. Nous y définirons la santé, la santé publique, la santé mentale et les soins en santé mentale.

Dans une étude statistique, il est important d'avoir un maximum d'informations sur la façon dont les données ont été récoltées. Nous consacrerons donc un chapitre à l'explication de la manière dont l'OWS a récolté ces données avec l'aide de chaque service de santé mentale wallon. Dans cette partie, les différentes étapes effectuées, dans le cadre de ce mémoire, pour passer des fichiers de base à ceux utilisés par la suite seront également développées.

Une fois le contexte clairement établi et la mise en place des données expliquée, une dernière partie sera encore développée avant d'analyser ces données. En effet, nous consacrerons le troisième chapitre aux données manquantes. Cette partie sera particulièrement axée sur les différents services ayant le plus de données manquantes, car cela était une demande de l'OWS. Nous observerons la quantité de données de ce type dans chaque variable, mais également la distribution de certaines données références au sein des données restantes. Ce chapitre permettra de bien fixer la qualité de la base de données.

Nous pourrons alors démarrer l'étude statistique qui sera développée en quatre chapitres. Dans un premier temps, deux chapitres analyseront respectivement le profil sociodémographique et le profil médical des nouveaux patients. Ceux-ci seront descriptifs et développeront chaque variable une à une, car le groupe de travail était intéressé par une description précise des données. Ensuite, ces deux types de données seront considérées simultanément, afin de faire des tests de dépendances entre des facteurs sociodémographiques et médicaux. De simples corrélations étaient demandées par le groupe de travail, mais nous avons décidé de les appuyer avec des tests d'hypothèses. Enfin, avant de conclure, le dernier chapitre tentera de réaliser une classification des patients.

Afin de vous permettre de mieux suivre la lecture, une annexe en fin de ce document contient un résumé du contenu de chaque variable. Normalement, tout est expliqué au fur et à mesure, mais n'hésitez pas à vous référer à cette annexe si nécessaire.

Chapitre 1

Contexte

Le contexte de ce mémoire fait partie d'une branche qui n'est pas nécessairement connue de tous : la santé publique. La santé mentale peut également sembler floue aux yeux de certains. Nous allons donc développer dans ce chapitre les notions qui nous paraissent importantes afin de comprendre le milieu dans lequel l'étude statistique est opérée, ainsi que l'importance de faire des études de ce type pour quantifier l'état de santé d'une population.

Avant d'expliquer en détails ce que sont la santé publique et la santé mentale, nous allons commencer par discuter de la santé en général. Nous pouvons avoir un premier aperçu de ses dimensions grâce à ces quelques définitions [20] :

- la santé est l'"état de celui dont les fonctions ne sont troublées par aucune maladie : être en bonne santé" ;
- la santé publique est "l'ensemble des actions et prescriptions de l'administration, relatives à la protection de la santé des citoyens" ;
- la santé mentale est "l'équilibre de la personnalité, la maitrise de ses moyens intellectuels".

Ce chapitre a pour but de bien fixer le cadre dans lequel va se situer ce mémoire. Nous aborderons d'abord de la santé en général. Ensuite, nous définirons la santé publique. Nous pourrons alors passer à la santé mentale et, pour terminer, nous parcourrons les différents soins proposés en santé publique.

1.1 La santé

Cette section se basera principalement sur le cours d' "introduction générale à la santé publique" des professeurs Gillet et Porignon donné à l'université de Liège en 2009-2010 [22]. Nous allons analyser plusieurs définitions de la santé, élaborées par différents auteurs .

Pour l'Organisation Mondiale de la Santé (OMS), "La santé est un état de complet bien-être physique, mental et social et ne consiste pas seulement en l'absence de maladie ou infirmité" (définition faite en 1948 par l'OMS de [22]). Il ne faut pas confondre la santé et la médecine. Quand nous parlons de la santé des citoyens, nous devons prendre en compte l'ensemble des facteurs qui affectent le bien-être de la population concernée, que ce soit au niveau physique, mental ou social. Bien sûr, le fait d'être malade ou non influence le bien-être physique et mental, mais beaucoup d'autres variables entrent également en ligne de compte.

"La santé n'est pas l'absence de maladie mais la conquête persévérante et lucide d'un équilibre de l'homme avec ce qui l'entoure" (J. Romains (Knock) [22]). La santé étant une "conquête", elle demande un effort de la part de l'individu concerné. Et la conquête étant "persévérante", cet effort doit être constant. De plus, l'individu est conscient des variables qu'il peut modifier ou non. C'est pour cela que la conquête est dite "lucide".

Nous citerons encore une dernière définition de la santé. "La santé est le niveau d'autonomie avec lequel l'individu adapte son état interne aux conditions de l'environnement tout en s'engageant dans le changement de ces conditions pour rendre son adaptation plus agréable et plus effective" (I. Illich [22]). La santé est donc, ici, une capacité d'adaptation. L'autonomie est, dans notre cas, considérée comme la capacité à prendre de bonnes décisions relatives à la santé. Cette définition est intéressante, car elle met en évidence le but de la santé publique. En effet, comme nous allons le voir dans la section suivante, les actions de santé publique ont pour but de quantifier et améliorer le niveau d'autonomie de la population.

Remarquons que toutes ces définitions montrent que les facteurs pouvant déterminer l'état de santé d'un individu sont nombreux et peuvent être d'ordre individuel ou collectif. Ainsi, nous avons les déterminants purement individuels, puis nous distinguons trois types de déterminants collectifs : les milieux de vie reprenant les influences sociales directement liées à la personne ; les différents systèmes comprenant les conditions de vie et de travail ; et le contexte global concernant les conditions générales socioéconomiques, culturelles et environnementales.

Nous avons donc des déterminants concernant des collectivités de plus en plus étendues. Chaque déterminant a une influence différente sur notre bien-être. Nous pouvons les représenter comme à la FIGURE 1.1.

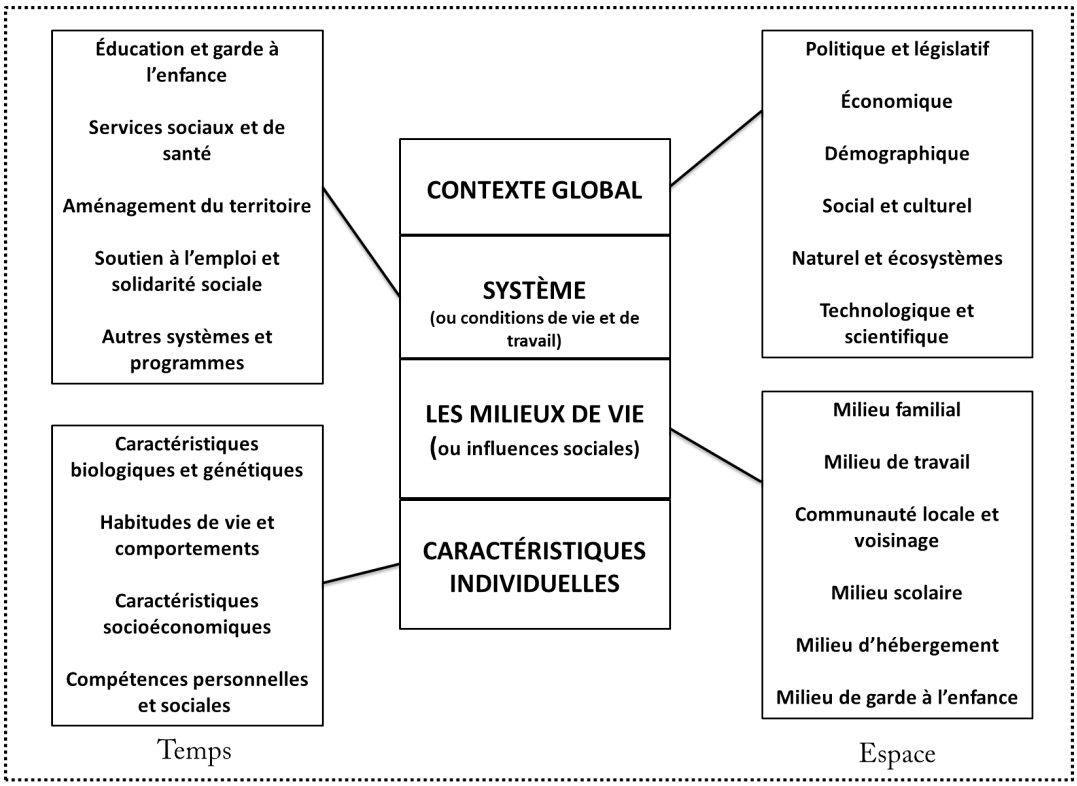


FIGURE 1.1 – Les déterminants de la santé (basé sur [22] et [21])

La FIGURE 1.1 est construite de sorte que, plus nous descendons, moins nous avons de personnes concernées par ces déterminants. En effet, les caractéristiques individuelles sont différentes pour chaque personne, les milieux de vie sont identiques pour des petits groupes, les systèmes englobent plusieurs de ces groupes et enfin, le contexte global concerne la collectivité entière. Les pointillés autour de l'ensemble indiquent que les différents champs interagissent et sont à prendre en compte ensemble.

De plus, le temps et l'espace doivent être pris en compte, car ils influencent ces facteurs. En effet, avec le temps, les mentalités des individus évoluent et le bien-être n'est plus nécessairement défini par les mêmes déterminants. Et même si c'était le cas, chaque déterminant n'a peut-être plus le même poids dans la santé de l'individu. L'espace est également indispensable, car l'endroit où l'on vit peut influencer notre bien-être. Lors de notre analyse, ces deux informations sont indispensables afin de pouvoir interpréter l'évolution au cours du temps et déterminer si certains lieux sont globalement mieux que d'autres ou si les habitants de cet endroit sont plus sensibles à certains facteurs.

Nous observons à présent chaque champ un à un afin de mieux cerner ce qu'ils englobent. Nous allons procéder en commençant par le bas du graphique, car nous sommes plus directement influencés par ce qui est plus proche de nous. Premièrement, les caractéristiques individuelles concernent tout ce qui influence le bien-être de la personne et qui lui est propre. Les individus peuvent réagir différemment face à une même situation

et ressentir les choses autrement. Ensuite, les milieux de vie correspondent à la sphère de la vie quotidienne de l'individu : son milieu social, son travail, etc. Ils imposent des conditions de vie aux niveaux matériel et social. Troisièmement, les systèmes, qui sont influencés par les choix politiques et les valeurs générales de la société, sont classés en cinq catégories. Enfin, le contexte global comprend l'organisation générale de la société concernée. Chaque être humain vit dans un environnement macroscopique particulier comprenant six grandes branches principales, comme on peut le voir sur la FIGURE 1.

Comme l'explique Mr Mpunikiye [21], les différents déterminants peuvent avoir des influences positives ou négatives sur le bien-être social, ainsi que sur la santé. Ces déterminants interagissent, mais ont également des composants indépendants (par exemple, le déterminant "sexe" sera peut-être indépendant des contextes). Ils ont, par conséquent, la possibilité de se neutraliser ou de se renforcer.

Il est très important d'évaluer l'état de santé global de la population, afin d'éventuellement pouvoir orienter les décisions politiques et proposer, si nécessaire, d'autres mesures à prendre en terme de santé publique.

1.2 La santé publique

Cette section sera également grandement inspirée par le cours d' "introduction générale à la santé publique" des professeurs Gillet et Porignon dispensé à l'université de Liège en 2009-2010 [22].

En 1951, C.E. Turner a écrit que "la santé publique, en même temps qu'une science, est l'art de prévenir les maladies, de prolonger la vie, de promouvoir la santé et de favoriser un rendement équilibré par des efforts organisés au niveau des collectivités [22]". Ce sont donc des actions menées à un niveau global de la société, mais qui ont pour but d'améliorer la santé de chaque individu vivant dans cette société. En 1996, l'Association Burkinabe de la Santé Publique (ABSP) définit deux dimensions de la santé publique : la discipline scientifique et la pratique professionnelle. Nous allons ici mentionner et commenter ces deux définitions.

> *La santé publique, en tant que discipline scientifique, couvre l'ensemble des activités visant à promouvoir, maintenir ou rétablir la santé des individus dans la communauté. Pour ce faire, elle met en oeuvre une approche globale, rationnelle et participative des problèmes de santé et fait usage d'un ensemble de méthodes appropriées dans un cadre pluridisciplinaire* (ABSP, 1996 dans [22]).

La santé publique définie comme une science, implique une certaine rigueur. Remarquons qu'ici, les approches de la santé publique sont "rationnelles" et "participatives", elles sont donc réfléchies afin d'être envisageables et nécessitent un minimum de bonne

volonté de la part des personnes concernées. De plus, afin d'arriver à ces objectifs, la santé publique suit "des méthodes appropriées". Elle ne fait donc pas n'importe quoi n'importe comment, mais suit certaines démarches prédéfinies. Enfin, son cadre est "pluridisciplinaire", par conséquent, elle n'agit pas qu'à un endroit, mais dans plusieurs disciplines, telles que la prévention, les soins ambulatoires, l'hygiène, etc.

> *La santé publique, en tant que pratique professionnelle, a pour objet les politiques de santé. À ce titre, le rôle principal est de déterminer les besoins de santé de la population, d'analyser la demande de celle-ci et de contribuer à l'organisation des services et des programmes communautaires de santé destinés à assurer le contrôle d'un problème de santé particulier ou la prise en charge d'un public particulier* (ABSP, 1996 dans [22]).

La santé publique, en tant que pratique professionnelle, est mise en oeuvre tous les jours. Remarquons qu'elle évalue les besoins de santé de la collectivité. Elle agit donc au niveau des personnes, qu'elles se sentent dans le besoin ou non. Nous touchons ici à une différence importante entre le médecin et la santé publique. Le médecin ne sait agir que si la personne se présente au cabinet. Une personne qui n'en ressent pas le besoin ne sera pas dirigée par le médecin. Par contre, la santé publique estime les besoins et agit auprès de l'ensemble de la population. De plus, elle organise les différents services et programmes accessibles aux personnes qui ont besoin d'aide.

Comme expliqué dans le cours [22], le clinicien et le praticien de santé publique ont des rôles complètement différents. En effet, le clinicien s'occupe de la maladie d'un seul individu malade et le praticien en santé publique de la santé de la communauté. Pour ce faire, le clinicien analyse les symptômes du malade, tandis que le praticien en santé publique analyse la situation globale de la communauté dont il s'occupe à l'aide de différents taux relevés. Enfin, le clinicien prend des mesures individuelles et le praticien en santé publique des mesures collectives.

1.3 La santé mentale

Le concept de santé publique étant établi en général, concentrons-nous sur la santé mentale. Nous allons ici définir ce concept et, dans la section suivante, nous parlerons des différents soins en santé publique et en santé mentale.

Notre développement ci-dessous est principalement basé sur les sites des deux organisations suivantes : l'Association canadienne pour la santé mentale - Chaudière-Appalaches [5] et le centre des médias de l'OMS [6].

Tout d'abord, il est important de se rendre compte que, comme pour la santé en général, la santé mentale ne concerne pas uniquement le fait de souffrir ou non d'une

maladie mentale. Elle considère le bien-être mental en général et comprend, par conséquent, tout ce qui peut influencer cet état mental. Pour l'Organisation Mondiale de la Santé (OMS), "la santé mentale englobe la promotion du bien-être, la prévention des troubles mentaux, le traitement et la réadaptation des personnes atteintes de ces troubles." [7]

Ce bien-être mental est influencé par des facteurs environnementaux, biologiques, socioéconomiques, psychologiques, etc. Selon le contexte global dans lequel vit l'individu (scolarité, pauvreté, politique, ...), il est plus ou moins sujet, a priori, à des troubles de la santé mentale. En effet, celle-ci est influencée , par exemple, par un travail éprouvant, des pressions ou de la discrimination. Pour l'Organisation Mondiale de la Santé Mentale, "une personne en bonne santé mentale est une personne capable de s'adapter aux diverses situations de la vie, faites de frustrations et de joies, de moments difficiles à traverser ou de problèmes à résoudre" (OMS dans [5]). L'ACSM, après avoir analysé les différentes définitions de la santé mentale, considère que c'est "la façon dont une personne pense, se sent et agit dans la vie" [5].

Le document "Psychiatrie et santé mentale" [16] amène encore des éléments supplémentaires qui sont développés à partir d'ici. La recherche concernant la santé mentale est moins avancée que la recherche de la médecine. De plus, il n'existe encore aucun vaccin permettant de se protéger d'une maladie mentale. Remarquons que la santé mentale doit être considérée dans le cadre individuel et dans le cadre collectif. En effet, la santé mentale influence les relations sociales, les capacités d'exercer un métier, etc. La personne qui souffre de problèmes mentaux va peut-être s'isoler de la société, car ce genre de maladies empêche parfois les relations avec les autres et n'est pas compris par tous.

Le champ de la santé mentale est divisé en trois parties [16]. Tout d'abord, "la santé mentale positive" qui permet de se sentir bien et de s'épanouir, de faire face aux aléas de la vie. Ensuite, "la détresse psychologique réactionnelle" est causée par une situation de la vie quotidienne difficile à gérer qui entraîne alors un sentiment de désespoir, qui est cette "détresse psychologique réactionnelle". Enfin, "les troubles psychiatriques" sont plus sérieux et nécessitent un diagnostic afin de guider la personne vers des traitements thérapeutiques adéquats. Ils peuvent être de durée et de gravité variables.

Par exemple, une personne qui, suite à une situation éprouvante, fait une tentative de suicide peut encore être reprise dans la seconde catégorie, alors que, quelqu'un qui essaye de se suicider plusieurs fois (suicidaire chronique) souffre de réels troubles psychiatriques et entre dans la troisième catégorie.

Ce concept est vaste et, pour pouvoir l'analyser, il faut trouver des indicateurs observables qui nous permettent de quantifier l'état de santé mentale d'une population. Ceux que nous prendrons en compte dans le cadre de ce mémoire seront développés plus loin, lorsque nous présenterons le questionnaire rempli par les services de santé mentale.

1.4 Les soins en santé publique

Remarquons que la santé mentale fait partie intégrante de la santé publique. Les actions menées pour améliorer la santé publique et la santé mentale ne sont donc pas nécessairement distinctes. En effet, par exemple, des services de santé mentale sont inclus dans certains hôpitaux généraux. Lorsque nous mentionnerons des exemples de soins, nous considérerons principalement la santé mentale, puisque c'est cette partie de la santé publique qui va nous intéresser par la suite. Les services de santé publique sont classés en trois catégories également appelées "lignes" par les professionnels (première, deuxième et troisième).

La première ligne concerne les aides directement accessibles lorsqu'une personne veut trouver de l'assistance en santé. "Les services de première ligne représentent le point de contact de la population avec le réseau de la santé." [8]. Les plus répandus sont les médecins généralistes qui peuvent identifier un problème de santé physique ou mentale et, si nécessaire, rediriger la personne vers un praticien spécialisé. Les assistants sociaux, le CPAS, les crèches, le planning familial, les maisons de repos, etc. sont également des services de première ligne.

Remarquons que les personnes souffrant de troubles mentaux ne s'en rendent pas toujours compte et ne demandent alors pas d'aide de ce côté. En effet, la santé mentale reste tabou dans notre société. Elle apparait comme secondaire par rapport à la santé physique. Par exemple, certaines personnes qui vont voir leur généraliste parce qu'elles se sentent déprimées vont mettre en avant un « alibi » physique : "je me demande si je ne fais pas une chute de tension".

Les personnes présentes sur place comme les assistants sociaux, les infirmiers à domicile, etc. peuvent, elles, voir comment l'individu se comporte et détecter un éventuel trouble d'ordre mental. Malheureusement, à l'heure actuelle, les personnes se trouvant en première ligne ne sont pas ou ne se sentent pas toujours aptes à réagir correctement face à ce genre de situation. Certains médecins généralistes se sentent également démunis face à des problèmes mentaux.

La deuxième ligne prend en charge les personnes nécessitant un service plus spécifique que la première ligne. C'est la première ligne qui envoie les patients dans les services de deuxième ligne. Donc, quelqu'un qui a besoin d'aide a toujours, dans un premier temps, un contact avec la première ligne. Si la première ligne juge cela nécessaire, le patient est alors dirigé vers la deuxième ligne. Nous avons, par exemple, les hôpitaux en deuxième ligne et dans le cas de la psychiatrie, les services psychiatriques en hôpital général, les structures conventionnées INAMI, les initiatives d'habitations protégées, les hôpitaux psychiatriques et les maisons de Soins Psychiatriques [11].

Les mêmes services peuvent être présents dans différentes lignes. Par exemple, les services de santé mentale et les psychologues relèvent à la fois de la première et de la deuxième ligne. La deuxième ligne soutient donc la première ligne, lorsque cela s'avère nécessaire.

La troisième ligne soutient encore la deuxième ligne, si nécessaire. Ce sont des services surspécialisés qui sont présents notamment dans certains hôpitaux universitaires. "Ils s'adressent à des personnes ayant des problèmes de santé très complexes, dont la prévalence est faible, ou dont la complexité requiert une expertise qui ne peut être offerte par les services de deuxième ligne." [24].

Remarquons que la maladie mentale est plus délicate à guérir que la maladie physique. Tout d'abord parce qu'il est moins facile de faire un diagnostic et, ensuite, parce que prendre des médicaments ou subir une opération ne permet pas d'enlever définitivement le trouble ou la maladie.

Une autre manière de caractériser les soins est de les diviser en deux catégories distinctes : les soins résidentiels et les soins ambulatoires. Les soins résidentiels sont ceux qui nécessitent que la personne reste sur place un certain temps. Cela signifie, d'une part, que la personne ne vit plus dans son milieu "naturel" et, d'autre part, qu'il faut des services avec des lits, des infirmiers, des repas, des cuisiniers, etc. À l'inverse des soins résidentiels se trouvent les soins ambulatoires. Dans ce cas, la personne vit chez elle et peut se rendre à ses soins ou accueillir certains spécialistes chez elle. Cela ne permet pas de pouvoir observer la personne continuellement.

À l'heure actuelle, autant pour des raisons de budget que de bien-être collectif et individuel, on essaye de réduire les soins résidentiels et d'y recourir uniquement si cela est vraiment indispensable. Les soins ambulatoires prennent en charge un maximum de patients, tant que ceux-ci n'ont pas absolument besoin d'une aide résidentielle.

Une réforme des soins psychiatriques est en cours d'expérimentation. Elle vise à structurer le secteur psychiatrique en un réseau qui ferait le lien entre les structures résidentielles et le domicile du patient. Dans ce cadre, tous les services de soins en santé mentale sont appelés à faire évoluer leur rôle. D'autres mesures de liaisons entre différents types de soins et services sont également prévues.

Chapitre 2

Nettoyage des données récoltées

Après avoir compris ce qu'était la santé publique et le contexte de ces données, une étape indispensable et assez longue a été de comprendre la façon dont les fichiers étaient créés et de convertir ces bases de données en un fichier clair et prêt pour une étude statistique. Dans ce chapitre, nous expliquons les différentes étapes parcourues pour obtenir cette base de données nettoyée.

Remarquons que les services de santé mentale n'ont pas tous rendu leurs fichiers sous le même format. Certains étaient en "excel" et d'autres en "access". L'OWS a déjà rassemblé les fichiers "excel" en ajoutant une colonne "service", afin de savoir quel patient était dans quel service. Ils ont décidé de faire, par année, deux fichiers : un premier comprenant les patients adultes lors de leur première consultation et un second reprenant les enfants.

Dans le cadre de ce mémoire, les fichiers encodés en "access" ont été joints un par un au bon fichier "excel", en ajoutant le nom de l'établissement dans la bonne colonne. Pour ce faire, chaque dossier a d'abord été décompressé et ensuite, la base de données a dû être dé-fractionnée. Remarquons que certains services de santé mentale ont laissé les lignes enregistrées les années précédentes avant de rendre leurs fichiers. Par exemple, le fichier de 2009 contenait des patients enregistrés pour la première fois en 2008. Par conséquent, afin d'éviter les doublons, il a fallu être prudent lors des concaténations des différents fichiers.

Pour remédier entièrement à ce problème, les dossiers de chaque établissement ont été vérifiés individuellement, afin de pouvoir identifier les doublons et n'en conserver qu'un seul exemplaire. Bien entendu, il n'est pas impossible que deux patients aient exactement les mêmes réponses. Nous avons retiré les doublons uniquement lorsque les deux numéros d'identifiants correspondants coïncidaient au sein des données de l'établissement.

En faisant cela, nous nous sommes rendu compte que, même si les formulaires ont été remplis, a priori, pour les patients enregistrés entre 2008 et 2011, les fichiers contenaient également des dates antérieures à 2008. De la même manière, certains patients arrivés

en 2012 se trouvaient déjà dans le fichier 2011. Nous ne considérerons, dans le cadre de ce mémoire, que les données des années 2008 à 2011.

De plus, certains services de santé mentale apparaissaient à différents endroits sous différents noms, comme par exemple "Charleroi Tramétis" et "Charleroi Grand Rue". Il a donc fallu identifier les services concernés afin d'uniformiser leurs identifications. Pour cela, plusieurs techniques ont été utilisées. D'abord, Véronique Tellier nous a fourni une liste des services de santé mentale agréés et une liste un peu plus récente se trouve sur le site de l'Institut Wallon pour la Santé Mentale (noté IWSM) [10] [1]. Ces deux documents se trouvent en annexe. Ensuite, lorsque le nom dans la colonne "service" ne paraissait nulle part ailleurs, nous avons effectué des recherches sur internet. Ainsi, nous avons pu faire le lien entre "Tramétis" et "Grand Rue". En effet, à Charleroi, le nom du service situé Grand Rue porte le nom "Tramétis". Enfin, quand cela s'avérait nécessaire, nous avons ouvert les fichiers de base de l'établissement, afin d'y détecter des similitudes avec d'autres fichiers.

Ce processus étant terminé, chaque fichier est désormais cohérent et prêt pour démarrer l'étude statistique. En effet, les fichiers de la forme "adultes*date*" possèdent bien uniquement des patients enregistrés dans l'année "*date*". De plus, chaque service de santé mentale apparaissant dans plusieurs années n'a désormais plus qu'une seule identification. Les 84 services présents dans les fichiers ont été insérés dans une liste. Cette liste est représentée à la TABLE 7.3, se trouvant en annexe.

Remarquons que certains services ont été créés en 2009 ou 2010. Nous n'aurons, par conséquent, aucune donnée pour ces services dans le fichier 2008. De plus, les données manquantes sont encodées comme "DM". Elles sont assez présentes dans certaines variables, car peu de services ont utilisé le code signifiant "Inconnu" (ou alors, la donnée a été perdue à un moment du processus, au cours des concaténations ou des transferts). Pour certaines données, telles que, par exemple, la source de revenu, il semble logique que le praticien ne connaisse parfois pas encore la réponse. Par conséquent, lorsque nous avons beaucoup de données manquantes et pas d'"Inconnus", il est raisonnable de considérer qu'une partie de ces "DM" sont, en fait, des données non connues. Cela est probable étant donné que les "DM" ont été insérées dans toutes les cases vides du fichier.

Au niveau de la composition du fichier, il est nécessaire de savoir que le questionnaire a dû être complété après seulement quelques rendez-vous. Donc, certaines données plus personnelles n'avaient pas encore été abordées lors des consultations. Ceci constitue une autre justification possible du nombre de données manquantes.

Après une première analyse des données manquantes destinées à valider la base de données, nous nous sommes rendu compte que certains services de santé mentale (ou SSM) ne possédaient aucune réponse pour plusieurs variables. Pour certains établissements, 100% des données étaient manquantes dans certaines colonnes. En abordant ce

1. la plupart des activités de l'IWSM sont reprises depuis le 01.01.12 par l'asbl CRéSaM (Centre de référence en santé mentale), reconnu et subventionné par la Région wallonne.

problème lors d'une réunion de groupe à l'OWS (le groupe est composé de Véronique Tellier, Dominique Dubourg, Marie Lambert, Christiane Bontemps, Anouck Billiet et Hugues Reyniers), nous nous sommes aperçus que cela concernait tous les services provinciaux d'une seule province. Après quelques recherches, nous avons trouvé l'erreur. Á partir d'une certaine variable, toutes les données étaient décalées d'une colonne pour ces services. Par exemple, pour la langue maternelle, il n'y avait que 4 modalités car ce sont celles de la variable "Nationalité". Les réponses concernant véritablement la langue maternelle se trouvaient dans la colonne de droite.

J'ai alors fait au mieux pour récupérer les données concernées mais, lorsque cela causait des confusions possibles, j'ai préféré encoder "ErreurEncodage" que de mettre des valeurs peut-être inexactes. Cela ne s'est généralement produit que dans les derniers choix des questions pouvant admettre plusieurs réponses. Par exemple, pour la variable "prises en charge antérieures", nous avons la prise en charge principale correctement encodée, mais pas la deuxième. Dans cette deuxième colonne, tous les services de cette province possèdent "ErreurEncodage".

Pour terminer, les données étaient sous forme de codes, telles que par exemple le sexe qui était 1 si le patient est un homme ou 2 pour une femme. Afin de faciliter l'analyse par la suite, nous avons remis chaque colonne en qualitatif. De cette manière, on voit plus facilement la valeur de chaque donnée.

Pour les analyses futures, il est utile d'avoir certaines données sous une autre forme. Par exemple, la date de naissance du patient ne nous intéresse pas directement, mais nous permet de calculer l'âge à la première consultation. Pour cela, il suffit de soustraire l'année de naissance à l'année d'enregistrement. Nous obtenons donc une variable dont les données manquantes sont celles où soit l'année d'enregistrement, soit l'année de naissance faisait défaut. De plus, nous avons retiré les âges incohérents qui devaient provenir de fautes d'encodage dans une des deux variables, comme par exemple les âges négatifs ou les âges supérieurs à 120.

Pour certaines analyses, il nous sera utile d'avoir des classes d'âges. Nous les avons déterminées avec le groupe de travail à l'OWS. Les classes d'âge se trouvent à la Table 2.1.

Intervalle	Nom de la classe
[18,24]	Jeunes
[25,44]	Jeunes adultes
[45,59]	Adultes
[60,120]	PersAgees
DM et Incohérences	HorsLimites

Table 2.1 – Classes des âges

De la même manière, à l'aide du code postal des nouveaux patients, nous avons créé une colonne contenant la province de leur domicile.

Une fois le nettoyage terminé, cette base de données contient, au total, les informations de 41 916 personnes adultes ayant consultés un service de santé mentale wallon pour la première fois entre 2008 et 2011.

A l'aide de la FIGURE 2.1, nous pouvons observer la répartition du nombre de patients enregistrés chaque année. Remarquons que, pour trois lignes de la base de données, nous n'avons pas l'année d'enregistrement. Nous avons encodé cette information en mettant dans la colonne "AnnéeEnregistrement" un "0".

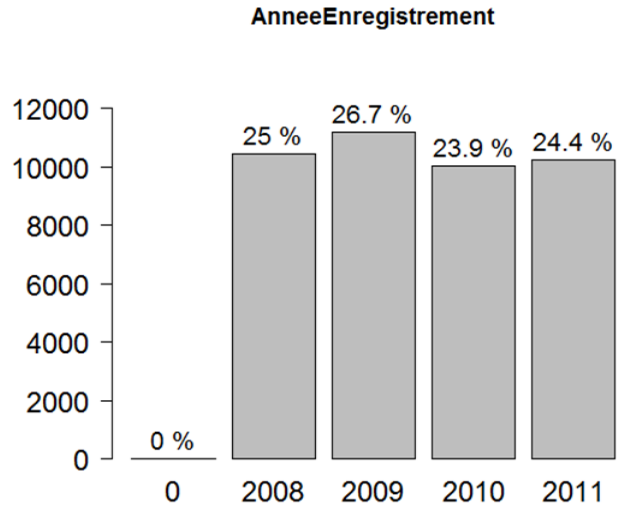


FIGURE 2.1 – Nombre de personnes par année d'enregistrement

Nous voyons que nous avons plus de nouveaux patients adultes en 2009 et moins en 2010. Il faut rester prudent sur le fait que cela ne signifie pas nécessairement qu'il y a eu plus de nouveaux patients cette année-là dans les services. En effet, nous n'avons pas la certitude que tous les patients ont été encodés, ni que toutes les données ont été rendues. De plus, il est possible qu'un service ait des patients "habituels" tellement nombreux qu'il ne soit plus en mesure d'accueillir, cette année-là, de nouveaux cas.

Au niveau du nombre de patients par service, le minimum est de 2 et le maximum de 6975. Ces deux valeurs sont extrêmes, mais cela peut s'expliquer facilement. En effet, le service avec seulement deux patients est, en fait, un service spécialisé pour les enfants. Par conséquent, il n'a reçu que très rarement des patients adultes. Pour le 6975, ce sont les "AIGS" (Association Interrégionale de Guidance et de Santé) qui sont un ensemble de services présents à divers endroits, mais qui ont été repris dans le même nom d'établissement dans les données. Nous avons donc beaucoup de patients dans les AIGS, puisque plusieurs services sont réunis dedans. Si nous observons la moyenne, nous obtenons 499 patients par SSM. Les AIGS, avec beaucoup plus de patients que

les autres, influencent beaucoup cette moyenne, donc nous pouvons calculer la médiane qui est 338, donc nettement inférieure à la moyenne. En conclusion, 50% des SSM ont reçu moins de 338 nouveaux patients de 2008 à 2011 et les autres 50% en ont reçu plus de 338. Pour avoir une idée plus précise de la répartition du nombre de nouveaux patients par SSM, nous avons fait une boîte à moustaches (ou boxplot), à la FIGURE 2.2.

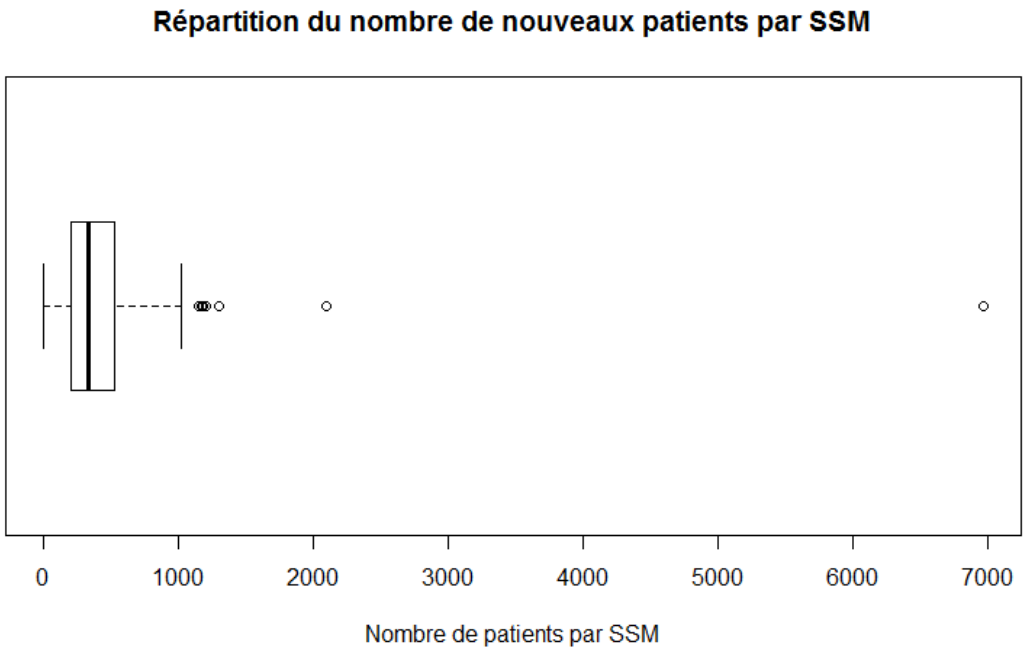


FIGURE 2.2 – Boxplot du nombre de nouveaux patients par SSM

Les moustaches de la boîte vont de 2 à 1024 patients et la boîte en elle-même démarre à 203.5 pour s'arrêter à 532.5. Nous observons la présence de quelques exceptions (appelées outliers) dépassant le maximum de la boîte. Cette boîte est créée de sorte à contenir à peu près 50% des données, en ayant 25% en-dessous et 25% au-dessus. Par conséquent, 50% des SSM ont eu un nombre de patients entre 203 et 532. Les quelques services possédant plus de 1024 patients sur les 4 ans sont considérés comme exceptionnels, ou outliers.

Chapitre 3

Analyse de données manquantes

La base de données, désormais prête à l'analyse, possède des données manquantes. Nous allons donc faire une analyse de représentativité des données. De plus, l'OWS nous a demandé de voir si certains services en possédaient plus que d'autres.

Certaines variables possèdent énormément de données manquantes (encodées "DM"). Les "DM" comprennent toutes les cases laissées vides dans la base de données. Malheureusement, cela inclut parfois également les cas où cette question n'était pas encore abordée lors des consultations. En effet, il existait un code pour indiquer que la donnée n'était pas encore connue par le praticien mais, malheureusement, il n'a pas été beaucoup utilisé. Par conséquent, lorsque le code "Inconnu" n'était utilisé que par une minorité de SSM, je les ai remplacés par des "DM".

Afin d'avoir une idée des proportions de données manquantes par variable, répertorions les dans un tableau. Nous allons, par la suite, considérer deux types de variables, celles qui définissent le "profil" des nouveaux patients, donc les données à caractère socio-démographique ; et celles qui concernent la consultation, telles que les raisons qui ont poussé le patient à se rendre dans un service. Pour les variables "profil", nous obtenons la table 3.1 :

Variable	Nationalité	Langue	EtatCivil	ModedeVie
DM	14%	7.7%	7.9%	7.1%

NiveauScol	CatégorieProf	SourceRevenus	Age	Province
39.4%	19.4%	12.9%	1.2%	3.4%

TABLE 3.1 – Proportions de données manquantes parmi les variables profils

Nous pouvons observer beaucoup de données manquantes dans les variables "NiveauScolaire" et "CatégorieProf". Cela peut s'expliquer par le fait que ces variables ne sont pas toujours abordées lors des premières consultations. Pour les autres variables, il y a moins de 15% de données manquantes. Dans l'ensemble, si les données présentes sont représentatives, ces données non définies n'influenceront pas les résultats. Ce tableau

nous permet de nous rendre compte que les données manquantes pourraient coïncider avec les informations non connues à cet instant-là.

Si nous regardons maintenant cela pour les variables concernant les consultations, nous obtenons la table 3.2. Nous pouvons y voir que beaucoup de données manquent dans certaines variables. Celles avec plus de 15% de données manquantes sont : "RéseauxProf", "PECAntérieure", "OrigineDémarche", "Ressources" et "ICD10".

Variable	TypeDossier	NatureDém	OrigineDém	Motifs	Demande
DM	0.4%	11.2%	29.6%	8.3%	8.3%

PECantérieure	Ressources	ICD10	PECproposée	RéseauxProf
46%	27.8%	17.1%	7%	54.4%

TABLE 3.2 – Proportions de DM parmi les variables concernant la consultation

Afin de voir si les données manquantes sont gênantes, nous allons voir si les données encodées sont représentatives. Pour ce faire, nous allons regarder si les proportions de diverses variables "représentatives" sont respectées. Nous ferons cette analyse uniquement pour les données ayant plus de 15% de "DM", car pour les autres il y a suffisamment de données encodées. Comme variables de référence, nous regarderons le sexe, l'année d'enregistrement et l'établissement, qui sont trois variables complétées pour presque tous les nouveaux patients.

Nous allons donc séparer cette section en trois sous-parties. Tout d'abord, nous regarderons le sexe, ensuite, l'année d'enregistrement, et pour terminer, les répartitions au sein des SSM.

Sexe

Pour commencer, nous allons nous focaliser sur le sexe. La FIGURE 3.1 nous montre plusieurs proportions d'hommes et de femmes. Dans un premier temps, nous avons celles au sein de toutes les données. Ensuite, nous avons cette répartition pour les données manquantes de chaque variable. Nous avons ajouté les proportions au sein de toutes les données afin d'avoir une référence, mais l'analyse de cette variable sera faite plus tard.

On peut voir sur ces histogrammes circulaires que les femmes restent majoritaires. De plus, les différents pourcentages de femmes sont compris entre 52.9% et 60.1%. On peut remarquer que les données les plus éloignées des proportions de départ sont celles qui avaient le moins de données manquantes. Puisque les proportions sont faites sur moins d'effectifs, il est difficile de savoir si cela pose un problème.

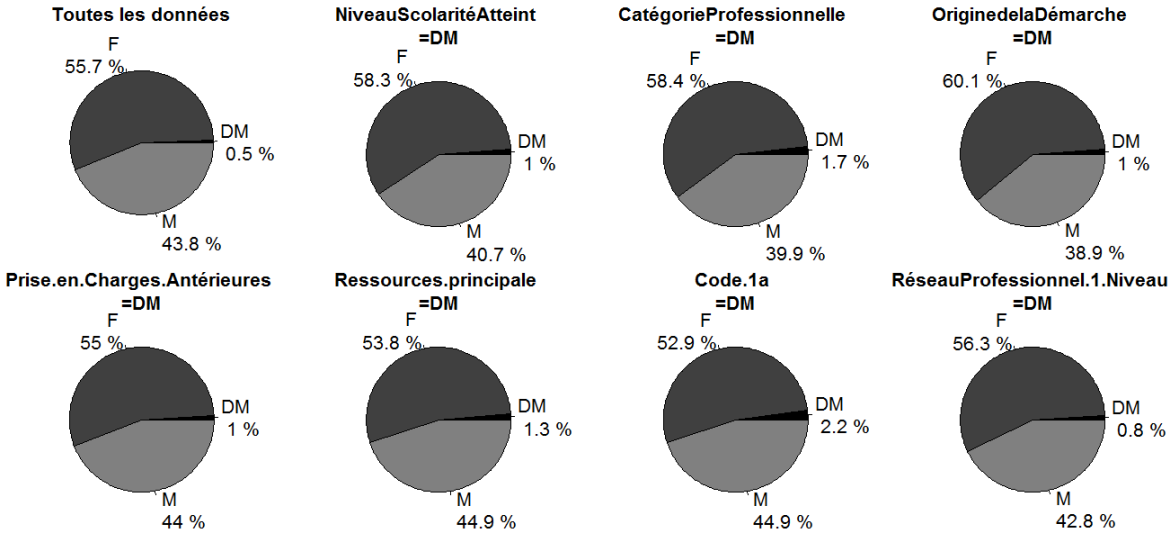


FIGURE 3.1 – Pourcentage de "Sexe" au sein des DM de chaque variable

Afin de se faire une idée statistique du fait que les proportions sont les mêmes ou différentes, nous procédons à un test d'hypothèses. Tester si les proportions sont égales revient à regarder si une dépendance est présente entre les données utilisées et la variable de référence, donc le sexe dans notre cas. Une explication précise du fonctionnement des tests d'hypothèses de dépendances sera faite dans le CHAPITRE 6 à la SECTION 6.1. Si les proportions sont égales, nous aurons une indépendance. Pour chaque variable, nous créerons donc un tableau de la forme de la TABLE 3.3.

	Dans toutes les données	Pour variable = DM
Hommes	XX	XX
Femmes	XX	XX

TABLE 3.3 – Tableau pour le test des proportions

Nous comparons donc toutes les proportions d'hommes et de femmes dans les DM à celles présentes dans l'ensemble des données. Nous prenons en considération qu'une p-valeur supérieure à 0.01 indique des proportions égales. L'explication de ce qu'est une p-valeur est également à la SECTION 6.1. Ici, nous ne nous intéressons qu'aux résultats. En faisant ces tests, nous obtenons les résultats de la TABLE 3.4 pour les premières variables. Nous constatons que, dans aucun cas, nous pouvons ne considérer les proportions d'hommes et de femmes égales dans l'ensemble des données et dans celles où ces variables sont manquantes.

	NiveauScolaritéAtteint	CatégorieProfessionnelle	OriginedelaDémarche
Résultat (P-valeur)	Différentes (0.000)	Différentes (0.000)	Différentes (0.000)

TABLE 3.4 – Test des proportions pour les variables = DM

Nous avons les résultats pour les autres données à la TABLE 3.5. Nous constatons que les proportions ne sont égales à celles de l'ensemble des données que pour les données manquantes des prises en charge antérieures et des réseaux professionnels. Ces variables sont celles qui avaient le plus de données manquantes.

PECAntérieures	RessourcesPrincipales	Code.1a	RéseauProfessionnel
Égales (0.2725)	Différentes (0.005)	Différentes (0.003)	Égales (0.052)

TABLE 3.5 – Test des proportions pour les variables = DM

Il parait donc judicieux de faire un nouveau graphique nous montrant la répartition des sexes au sein des données effectives. En effet, il n'est pas nécessaire que les données manquantes, mais bien que celles remplies, respectent la répartition d'hommes et de femmes pour ne pas être biaisées. Or, il est possible que le sexe des personnes dont les données ne sont pas définies ne respecte pas la répartition totale, alors que les données restantes en sont proches, surtout pour les variables avec peu de données manquantes.

Nous pouvons voir cette répartition à la FIGURE 3.2, qui nous indique que les pourcentages de femmes au sein des données complétées restent entre 53.8% et 56.4%. Cet intervalle semble grand. Cependant, dans l'ensemble des données, nous avons 55.7% de femmes. L'intervalle se situe donc bien autour de cette valeur. Nous allons, à nouveau, analyser ces répartitions plus statistiquement.

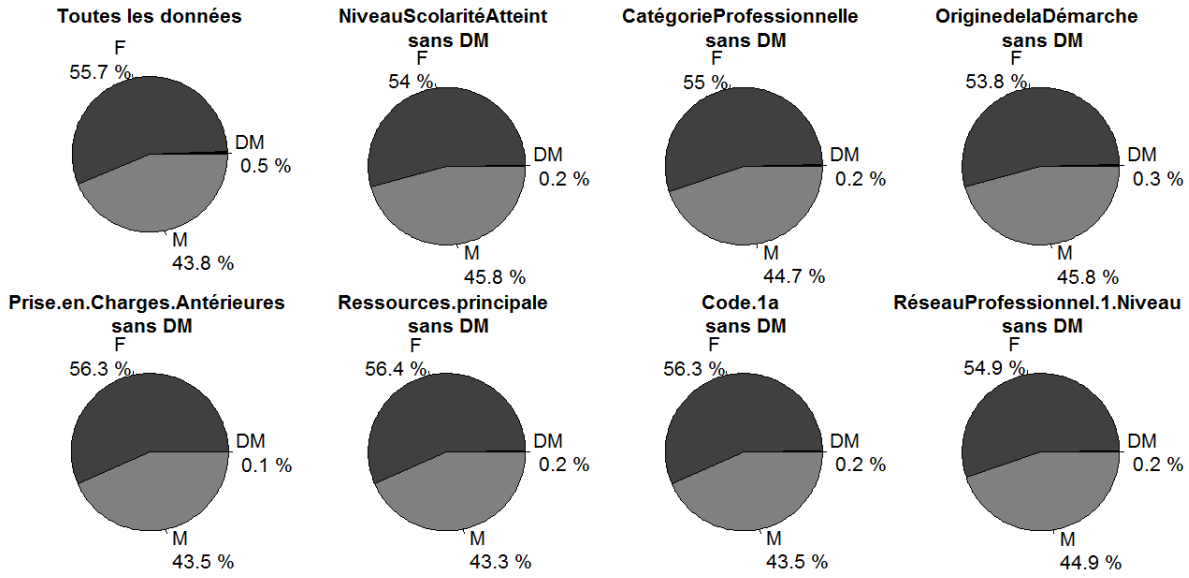


FIGURE 3.2 – Pourcentage de "Sexe" au sein des données sans DM de chaque variable

Si nous procédons à des tests d'égales proportions, nous obtenons pour les premières variables la TABLE 3.6. Nous observons que, pour le niveau de scolarité et l'origine de la démarche, les proportions semblent un peu trop éloignées. En effet, dans ces deux données, il semble y avoir moins de femmes par rapport à l'ensemble de l'échantillon. Cela est à garder en tête lorsque nous analyserons ces données.

	NiveauScolaritéAtteint	CatégorieProfessionnelle	OriginedelaDémarche
Résultat	Différentes	Égales	Différentes
(P-valeur)	(0.000)	(0.025)	(0.000)

TABLE 3.6 – Test des proportions pour les variables socio-économiques sans DM

Pour les autres variables, la TABLE 3.7 indique que les proportions peuvent être considérées comme égales. Par conséquent, pour ces données, nous pouvons considérer que l'échantillon des données remplies est bien représentatif en terme de sexe des patients.

PECAntérieures	RessourcesPrincipales	Code.1a	RéseauProfessionnel
Égales	Égales	Égales	Égales
(0.326)	(0.138)	(0.293)	(0.030)

TABLE 3.7 – Test des proportions pour les variables "consultation" sans DM

Au total de toutes les variables avec plus de 15% de données manquantes, seuls l'origine de la démarche et le niveau de scolarité ne semblent pas représentatifs au niveau de la proportion d'hommes et de femmes. En effet, légèrement moins de femmes sont présentes dans les réponses données pour ces deux questions.

Année d'enregistrement

La variable suivante à observer afin de se faire une idée de la représentativité des données encodées est l'année d'enregistrement. Nous pouvons voir la proportion de chaque année d'enregistrement pour les données manquantes à la FIGURE 3.3.

Nous pouvons voir que les proportions ne sont pas très similaires dans certains cas. Par exemple, pour la catégorie professionnelle, les années 2010 et 2011 sont surreprésentées. Les données manquantes ne représentant qu'une partie des patients encodés, il est intéressant, comme pour le sexe, de vérifier les proportions au sein des données restantes. Par exemple, la catégorie professionnelle possède des DM pour moins d'un cinquième des patients de la base de données. Donc, si les proportions ne sont pas fort semblables au sein des données manquantes, elles le seront peut-être quand même plus dans les données correctement encodées.

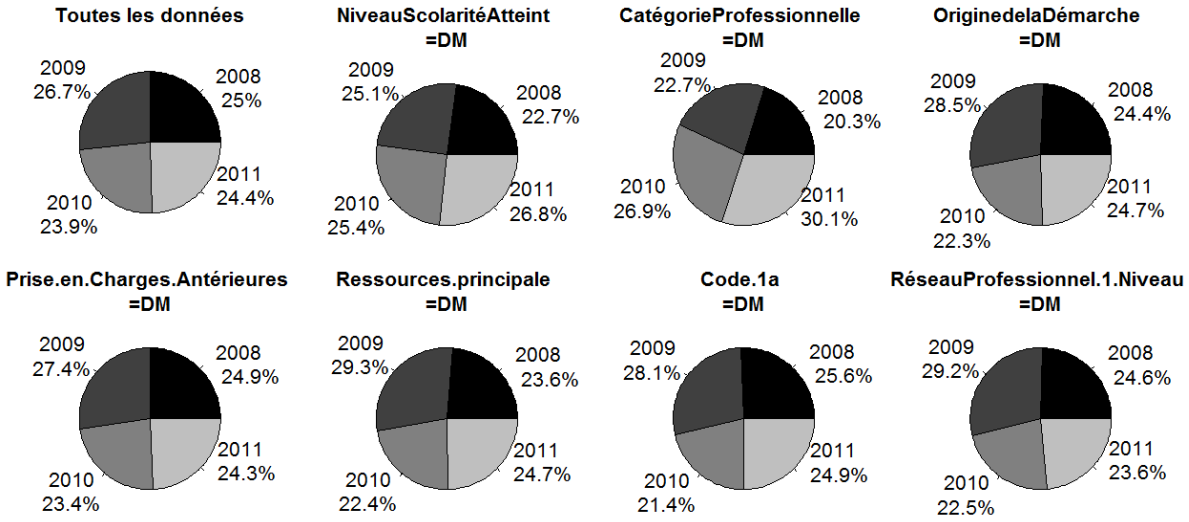


FIGURE 3.3 – Pourcentage de "AnneeEnregistrement" au sein des DM de chaque variable

Notons que si nous procédons aux tests d'égales proportions avec les données manquantes, nous obtenons que seule la variable concernant les données manquantes des prises en charge antérieures a des proportions égales à toutes les données.

Nous avons à la FIGURE 3.4 les années d'enregistrement des données non manquantes. Nous pouvons voir que, pour la catégorie professionnelle, les années d'enregistrement des données effectives sont plutôt bien réparties par rapport à l'ensemble des données.

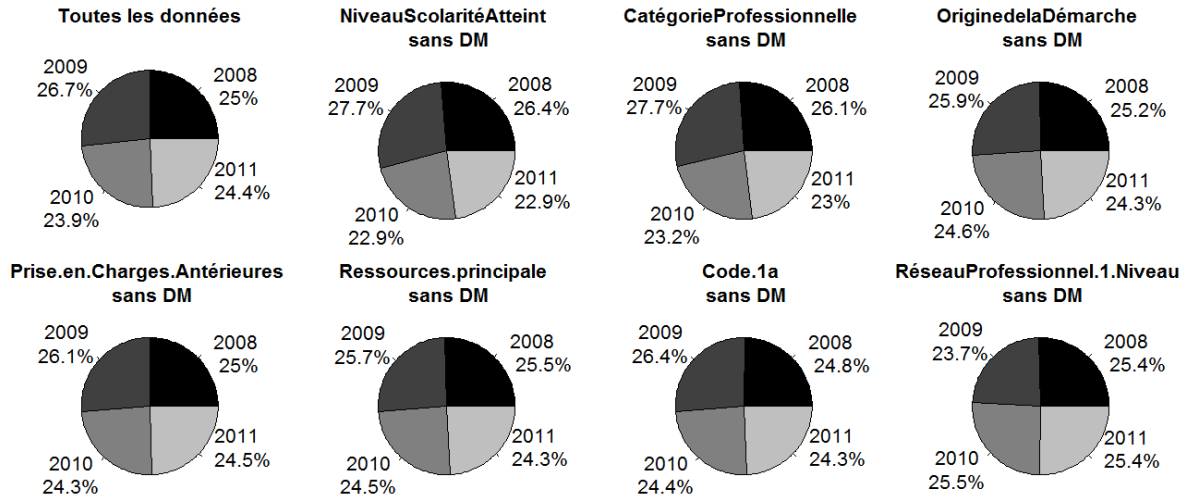


FIGURE 3.4 – Pourcentage de "AnneeEnregistrement" au sein des données sans DM de chaque variable

En procédant aux tests de proportions égales sur ces données, nous obtenons qu'il y a trois données pour lesquelles les proportions des années d'enregistrement ne sont pas similaires : le niveau de scolarité, la catégorie professionnelle et les réseaux professionnels. En observant la FIGURE 3.4, nous pouvons identifier les différences. En effet, pour le niveau de scolarité, nous avons proportionnellement plus de données en 2008 et 2009 et moins en 2010 et 2011 que dans l'ensemble des données. Nous constatons que cette variable a été de moins en moins remplie au fur et à mesure du temps. L'observation de la catégorie professionnelle donne des résultats similaires. Pour le réseau professionnel, les proportions en 2008 sont similaires, puis en 2009 nous avons moins de réponses, pour ensuite en avoir plus en 2010 et 2011. La différence de proportions de cette donnée est guidée par cette diminution en 2009.

SSM

Pour terminer l'analyse de représentativité, considérons les services de santé mentale. Dans ce cas-ci, il ne serait pas clair de représenter la proportion de chaque SSM dans les données manquantes. Cela est dû au fait qu'il y a beaucoup de services différents et que les patients ne sont pas distribués uniformément au sein des services (certains SSM reçoivent plus de nouveaux patients adultes que d'autres). Observons plutôt le pourcentage de données manquantes au sein de chaque service de santé mentale. De cette manière, nous pourrons identifier si tous les services ont le même taux de données manquantes, ou si celles-ci viennent plus de certains services. De plus, cette analyse est un feedback intéressant sur la façon dont les services remplissent ces données. Pour la suite, cela pourrait permettre de donner des pistes à certains SSM pour être plus efficaces lors de l'encodage et de la manipulation des données.

Pour faire les graphiques qui vont suivre, j'ai d'abord créé un vecteur reprenant, pour chaque SSM, le pourcentage de données manquantes dans la variable concernée. Ensuite, j'ai généré un histogramme de ces pourcentages, ainsi que la densité qui y serait associée.

Commençons par analyser les variables "profils" ayant au moins 15% de données manquantes. Nous pouvons voir à la FIGURE 3.5 les graphiques concernant la variable "NiveauScolaireAtteint". Nous observons que la forme ressemble approximativement à une gaussienne de moyenne proche des 20% et que très peu de services ont plus de 80% de données manquantes. De plus, moins de 15% des services ont plus de 60% de données manquantes dans cette variable. Cette courbe ne parait pas inquiétante, même si certains services semblent avoir plus de données manquantes que d'autres. Rappelons que certains services n'ont que 4 ou 5 patients adultes et que, par conséquent, une seule donnée manquante leur fait déjà 25% de "DM".

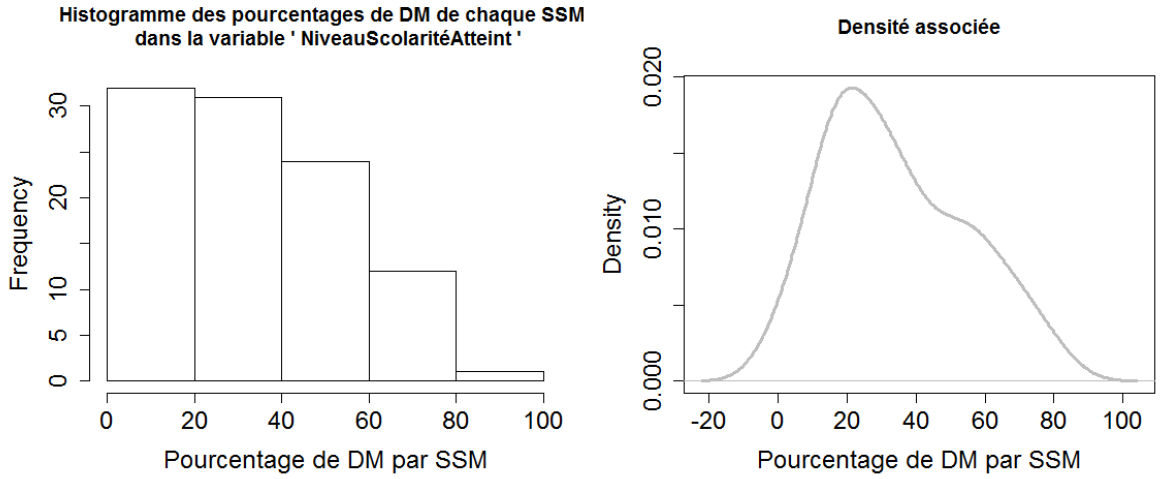


FIGURE 3.5 – Pourcentage de données manquantes dans "NiveauxScolaire" par SSM

Nous pouvons passer à la FIGURE 3.6 représentant les données manquantes dans la colonne "CatégorieProfessionnelle". Cette courbe est tout ce qu'il y a de plus rassurant. En effet, plus de 70% des services ont moins de 20% de données manquantes dans cette donnée et la densité consiste en un pic en moins de 10% avec une courbe très basse à partir de 40%. Nous pouvons donc, ici, considérer que les données manquantes sont bien réparties au sein des différents SSM.

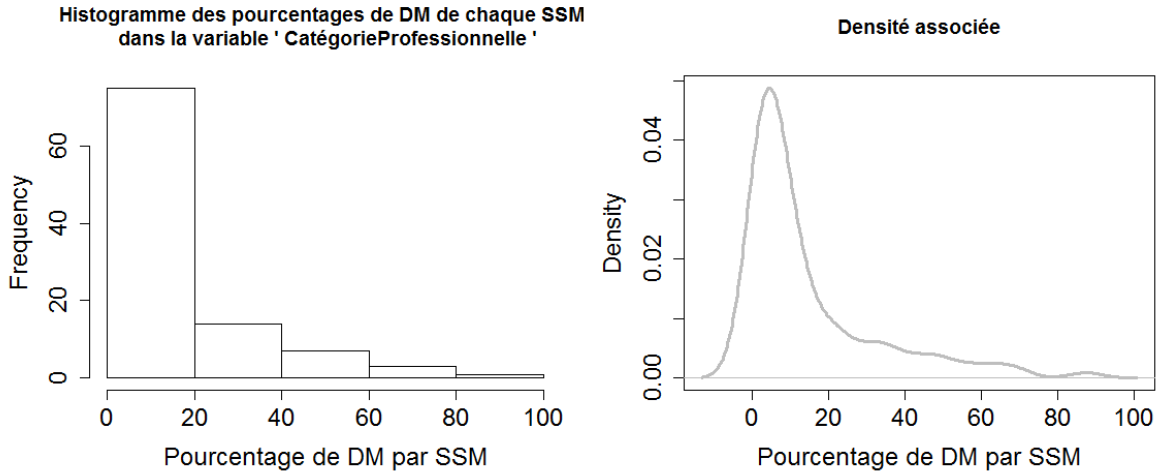


FIGURE 3.6 – Pourcentage de DM dans "CatégorieProfessionnelle" par SSM

Nous allons maintenant nous préoccuper des variables concernant la consultation. Tout d'abord, abordons l'origine de la démarche. La FIGURE 3.7 nous montre qu'on a plus de 60% des SSM qui ont moins de 20% de données manquantes dans cette colonne. Ensuite, on a bien une courbe fort décroissante, mais au dessus de 80% de DM, nous

avons un deuxième pic. Nous avons donc probablement quelques services pour lesquelles des problèmes d'encodage ou des erreurs de manipulations se sont produits.

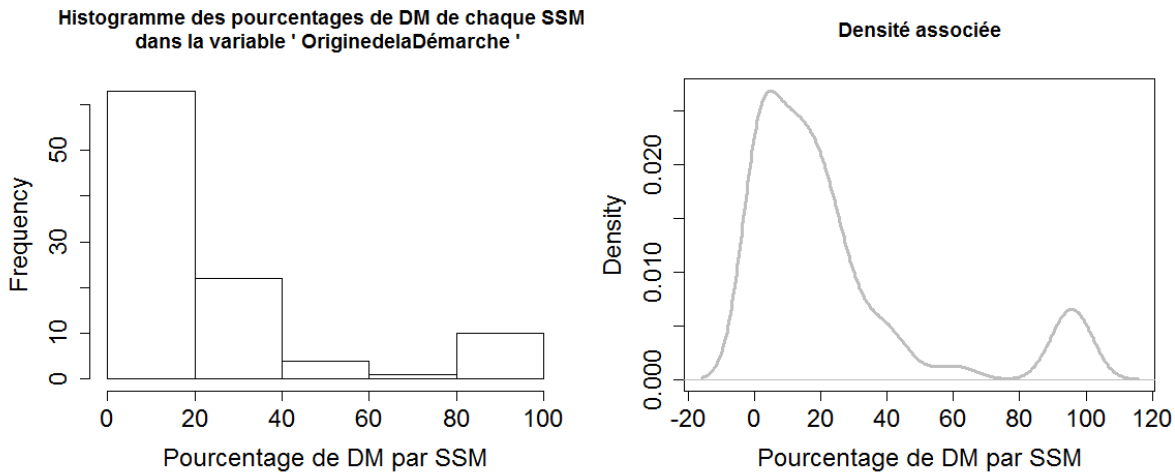


FIGURE 3.7 – Pourcentage de DM dans "OriginedelaDémarche" par SSM

Afin de permettre à l'OWS de savoir pour quels services nous avons eu ce souci, j'ai fait l'histogramme des services ayant plus de 80% de données manquantes. Afin d'assurer un minimum de confidentialité au niveau des services, j'ai introduit aléatoirement un chiffre à chaque service. La liste de correspondance entre ces chiffres et le nom des établissements ne sera fournie qu'à l'OWS qui pourra décider de qui y aura accès. Nous avons cet histogramme à la FIGURE 3.8.

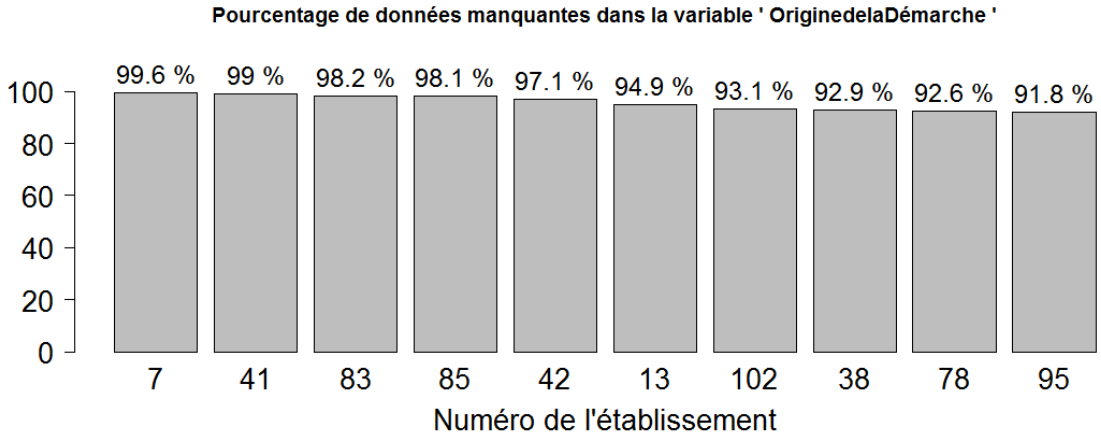


FIGURE 3.8 – Pourcentage de DM dans "OriginedelaDémarche" par SSM

Nous pouvons constater, à la FIGURE 3.8, que tous les services ayant plus de 80% de DM en ont en fait plus de 90%. Aucun service n'a entre 80 et 90% de données manquantes. Remarquons qu'en observant les correspondances, nous avons pu constater que toutes ces données manquantes proviennent des services provinciaux de la province du Hainaut. Il faudra donc garder en tête, lors de l'analyse de cette variable, que la province du Hainaut est sous-représentée. Par contre, pour les autres provinces, les données manquantes ne posent aucun problème.

Au niveau des prises en charge antérieures, les graphiques sont sur la FIGURE 3.9. Nous pouvons voir que les services sont fort présents dans tous les cas de figures de données manquantes. En effet, près de 30% des SSM ont entre 0 et 20% de données manquantes dans cette variable. La partie la moins représentée est celle entre 60 et 80% de données manquantes, mais concerne malgré tout plus de 10% des services.

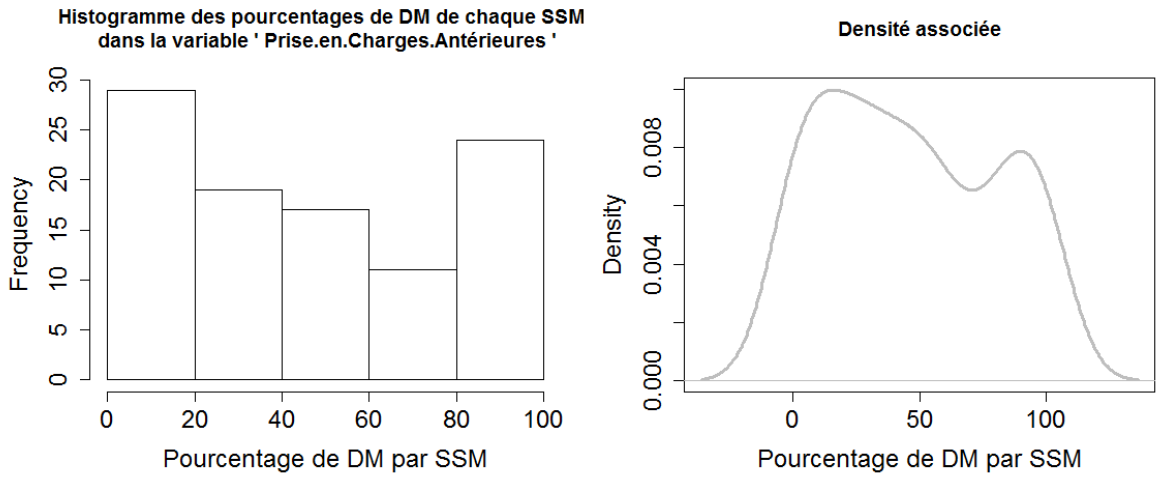


FIGURE 3.9 – Pourcentage de DM dans "PECAntérieures" par SSM

Remarquons qu'en analysant les différents pourcentages pour chaque SSM, nous avons pu remarquer que cette variable ne contient pas de problèmes spécifiques à la province du Hainaut. En effet, certains services de cette province font partie de ceux avec très peu de données manquantes dans cette variable. Afin de fournir à l'OWS les services les moins représentés dans cette variable, nous ajoutons un histogramme avec les SSM possédant plus de 90% de données manquantes. Cette illustration se trouve à la FIGURE 3.10. Nous constatons que les centres correspondant sont tous les "AIGS" (Association Interrégionale de Guidance et de Santé).

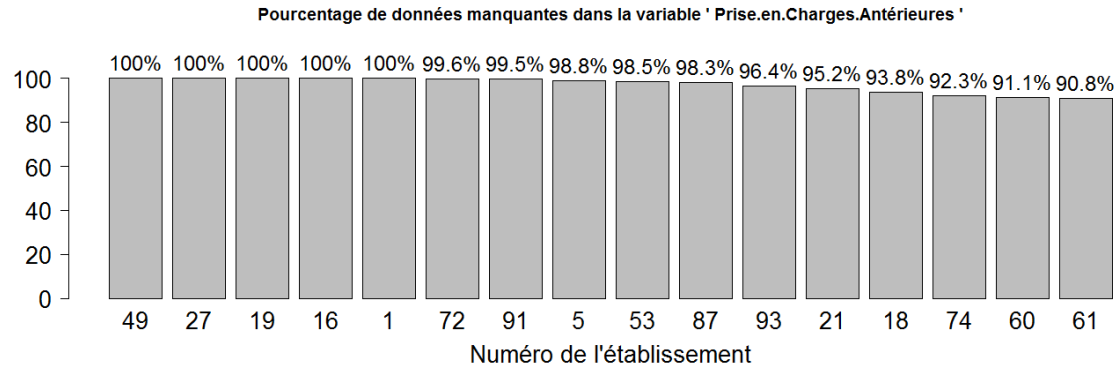


FIGURE 3.10 – Pourcentage de DM dans "PECAntérieures" par SSM

Pour les ressources, nous obtenons la FIGURE 3.11. Nous voyons directement que la majorité des services n'ont que très peu de données manquantes dans cette variable. En effet, plus de 50% des services ont entre 0 et 20% de données manquantes. Nous avons ensuite une courbe qui décroît assez vite. La courbe est, ici, également rassurante.

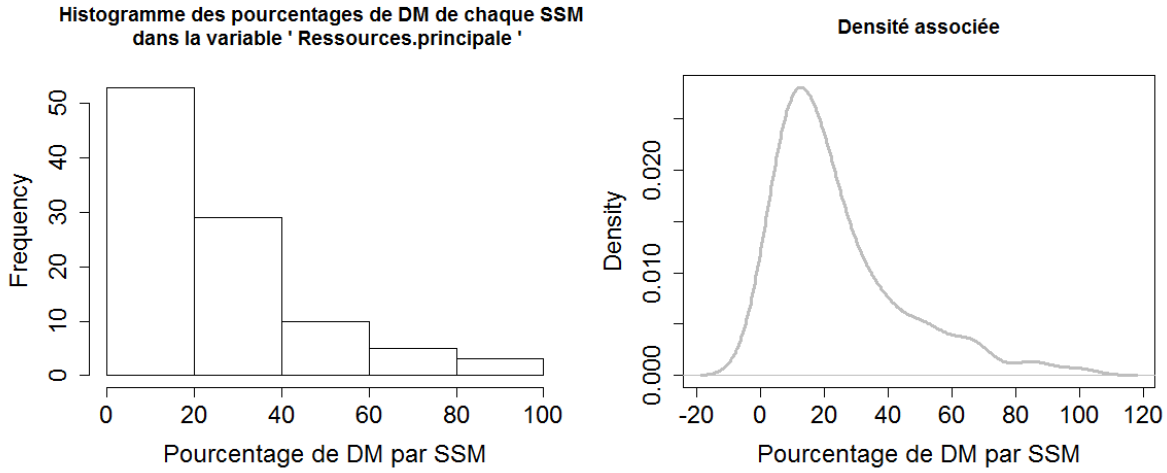


FIGURE 3.11 – Pourcentage de DM dans "Ressource" par SSM

Pour les codes ICD10, la FIGURE 3.12 nous indique que la plupart des SSM n'ont que très peu de données manquantes. Nous avons plus de 80% des SSM dont on a moins de 20% de données manquantes. Le pourcentage de données manquantes est élevé uniquement parce que quelques services possèdent beaucoup de données manquantes. Ces établissements ne correspondent pas à une même province. Seul l'établissement numéro 30 a presque 100% de données manquantes. Les autres en ont moins de 85%.

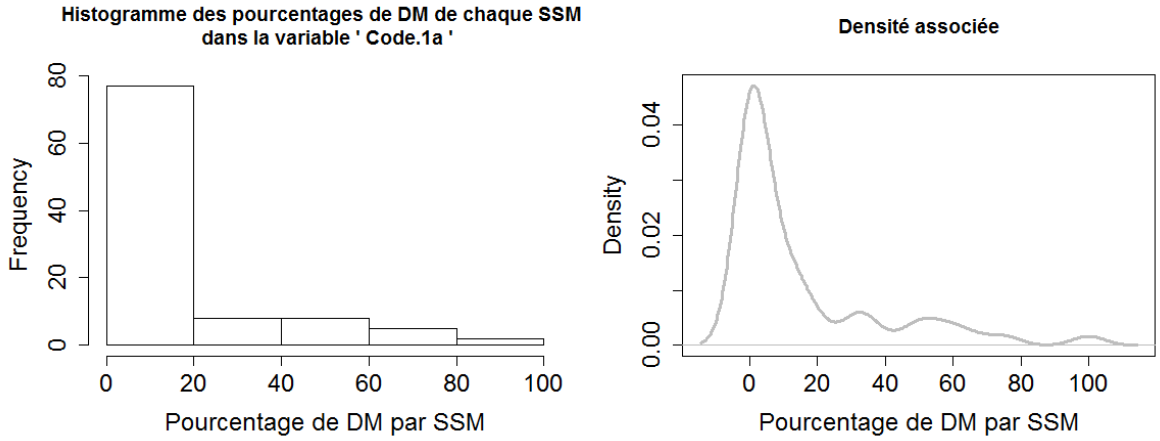


FIGURE 3.12 – Pourcentage de DM dans "CodeICD10" par SSM

La FIGURE 3.13 nous montre les données manquantes au sein des données "RéseauxProfessionnels". Nous pouvons voir que la plupart des établissements ont entre 40 et 60% de données manquantes dans cette variable. Il semble important ici de rappeler que l'information permettant de différencier les "DM" des "Inconnus" n'a pas toujours été utilisée et a peut-être été perdue dans d'autres cas. Par conséquent, ce grand nombre de données manquantes vient probablement du fait que cette donnée n'est pas toujours connue par le praticien après les premiers rendez-vous. Cette variable concernant les autres réseaux professionnels qui soutiennent le patient est susceptible de ne pas être déterminée dans la majorité des cas.

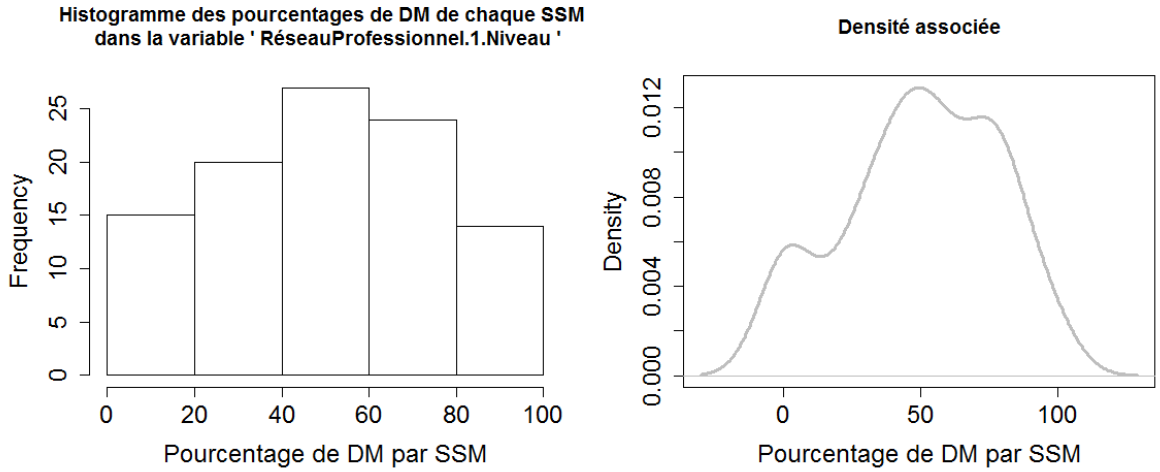


FIGURE 3.13 – Pourcentage de DM dans "ReseauxProf" par SSM

Conclusion

Nous pouvons conclure que les données manquantes ne posent globalement pas trop de problèmes en terme de représentativité des données. En effet, pour la plupart des variables, les données effectives ne sur-représentent pas un sexe ou une année. Cependant, il faut mentionner que le niveau de scolarité atteint et l'origine de la démarche sont remplis pour moins de femmes que la répartition des données complètes. De plus, au niveau des années d'enregistrement, quelques différences se font sentir au niveau du niveau de scolarité, de la catégorie professionnelle et des réseaux professionnels.

L'analyse des DM nous montre qu'il faut également rester prudent au niveau des services pour certaines variables. Cela est expliqué dans la section 'SSM' ci-dessus.

Chapitre 4

Profil socio-démographique

Le but de ce chapitre est de décrire les données socio-démographiques disponibles sur les nouveaux patients. Nous ferons des liens entre plusieurs variables de ce type lorsque cela aura été demandé par l'OWS. Ce chapitre illustre la répartition de chaque variable donnée par donnée, car cela était important pour le groupe de travail de se rendre compte des caractéristiques de la partie de la population qui consulte.

Pour établir le profil des nouveaux patients, il est important d'observer, dans un premier temps, leur sexe et leur âge. Si nous générons à l'aide du programme "R" un diagramme circulaire du sexe de toutes les personnes répertoriées entre 2008 et 2011, nous obtenons le graphique de gauche de la FIGURE 4.1. Ce diagramme nous montre que 55.7% des patients adultes sont des femmes. Afin de voir si cela signifie que les femmes sont plus concernées, il est nécessaire de comparer cette proportion avec celle de l'ensemble de la population wallonne. Pour cela, la base de données de 2008 nous a été fournie par l'OWS. Le diagramme circulaire correspondant est celui de droite de la FIGURE 4.1.

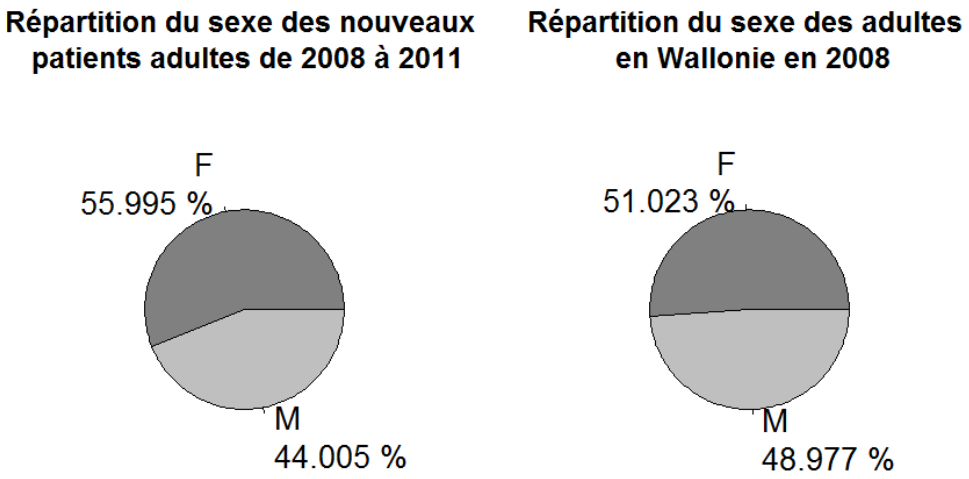


FIGURE 4.1 – Diagramme circulaire du sexe des patients adultes

Remarquons que les données disponibles pour toute la Wallonie concernent l'année 2008, qui est inclue dans les années dont nous avons les données des SSM. Nous constatons qu'au sein de la Wallonie entière, les femmes sont légèrement majoritaires. En effet, la population Wallonne de 2008 est composée de plus de 51% de femmes. Il y a donc plus de femmes, mais la disproportion n'est pas énorme. Par contre, dans nos données, nous avons également une majorité de femmes, mais la proportion est plus importante, puisque nous avons presque 56% de femmes dans la base de données. Il semble donc évident que les femmes sont plus vite amenées à se rendre dans des services de santé mentale. Cela se confirme par un test d'égales proportions ressortant une p-valeur inférieure à 10^{-16}. Cela ne signifie pas nécessairement que les femmes sont plus touchées par des problèmes de santé mentale, mais simplement que les femmes sont plus vite amenées à se rendre pour la première fois dans un service.

Nous pouvons maintenant nous intéresser à l'âge. La représentation de cette donnée se trouve à la FIGURE 4.2. Nous pouvons observer deux pics : un plus prononcé autour de la quarantaine et un moins élevé autour de la vingtaine. Nous pouvons voir que ces pics ne sont pas présents au sein de la base de données reprenant toute la Wallonie en 2008. De plus, au delà de 45 ans, la courbe de notre récolte de données est beaucoup plus décroissante que celle de la Wallonie. Par conséquent, les moins de 45 ans sont plus susceptibles de se rendre dans un SSM, avec un pic autour de 20 et 40 ans, alors que pour les personnes de plus de 45 ans, il y a moins de chance de s'y rendre. Au delà de 65 ans, nos données sont déjà très basses et décroissent encore lentement, alors que l'ensemble de la Wallonie reste encore assez haut jusqu'à 90 ans. Ces graphiques nous permettent de voir qu'il est évident que la distribution des âges de la population complète et de celle qui fréquente les SSM est différente.

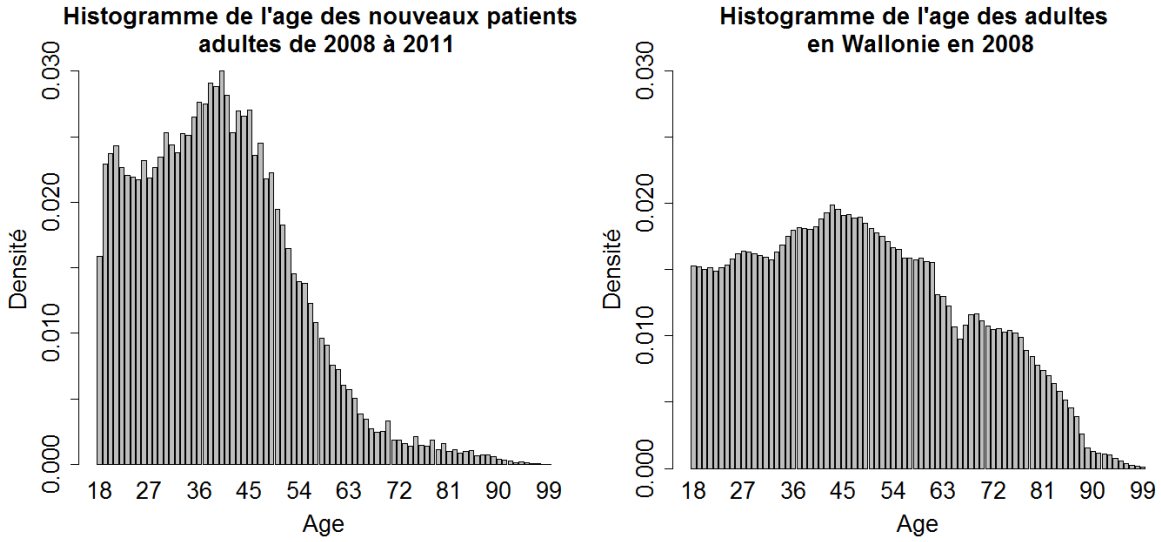


FIGURE 4.2 – Histogramme des âges

Avec l'aide du groupe de travail, nous avons pu identifier que le pic de la vingtaine concernait les patients qui doivent rentrer dans le monde du travail et sont à un moment charnière de leur vie, et que le pic de la quarantaine coïncide avec la "crise de la quarantaine".

Pour finaliser l'étude de l'âge, nous pouvons calculer les différentes moyennes et les écarts-types. Dans nos données, nous avons une moyenne de 39.4 ans et un écart-type de 13.63, alors que, pour la Wallonie, la moyenne est de 48.4 ans et l'écart-type de 18.47. Par conséquent, la population qui se rend dans les SSM a tendance à être plus jeune. En effet, la moyenne est plus petite, et l'écart-type également, donc la population est plus agglomérée autour de la moyenne.

Cette différence entre les moyennes peut être appuyée et vérifiée à l'aide d'un test d'hypothèses dont l'hypothèse de base est "la moyenne de nos données est celle de l'ensemble de la Wallonie". Pour pouvoir faire ce test, appelé "test t de student", il faut que les données suivent une courbe gaussienne (en cloche). La FIGURE 4.2 nous donne une première idée de la forme de la courbe, qui ressemble à une loi normale autour de 49 ans, mais avec un deuxième pic à 20 ans. Afin d'observer plus précisément si la courbe est gaussienne, nous pouvons tracer un graphique quantiles-quantiles. En abscisse, nous avons les quantiles d'une loi normale théorique et en ordonnée ceux de nos données. Par conséquent, si l'échantillon suit exactement une loi normale, la courbe doit être une droite. Nous avons ce graphique à la FIGURE 4.3. Entre 23 et 60 ans, nous avons bien une droite. Donc, dans cet intervalle, la courbe est gaussienne. Nous avons bien une loi normale autour de la moyenne et c'est sur les extrémités que nous nous en éloignons.

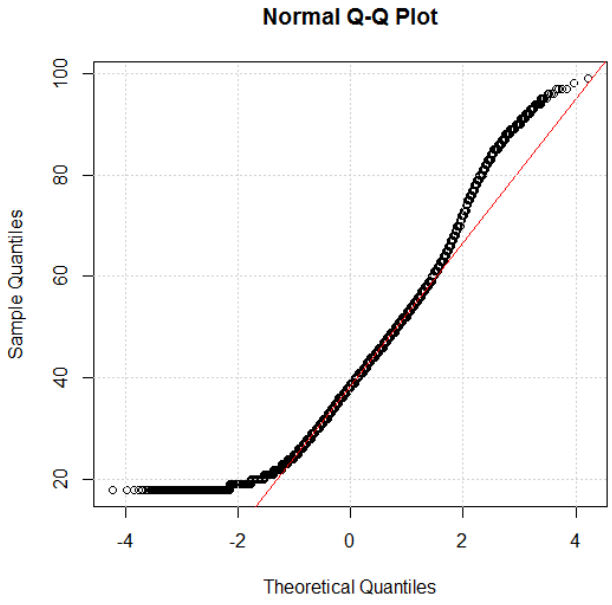


FIGURE 4.3 – Graphique quantiles-quantiles des âges des nouveaux patients

La ligne rouge correspond à la droite qui relie le premier et le troisième quartile. La partie gauche du graphique s'explique par le fait qu'il y a beaucoup de patients entre 18 et 22 ans, comme nous l'avons vu sur l'histogramme. Cela implique que les premiers quantiles de notre base de données sont tous entre 18 et 22 ans. Une fois ce pic passé, la courbe est similaire à une droite, et même une fois que les points passent au dessus de la droite rouge, ils se stabilisent rapidement en une droite parallèle à la rouge. Ce petit écart nous indique que nous avons moins de données aux alentours des 80 ans que celles que nous aurions observées avec une partie gaussienne. Nous avons tracé la densité de nos données, ainsi que celle correspondant à une loi normale de moyenne et d'écart type égaux à ceux de notre échantillon à la FIGURE 4.4. Nous confirmons que c'est aux extrémités que la densité ne suit pas vraiment la gaussienne.

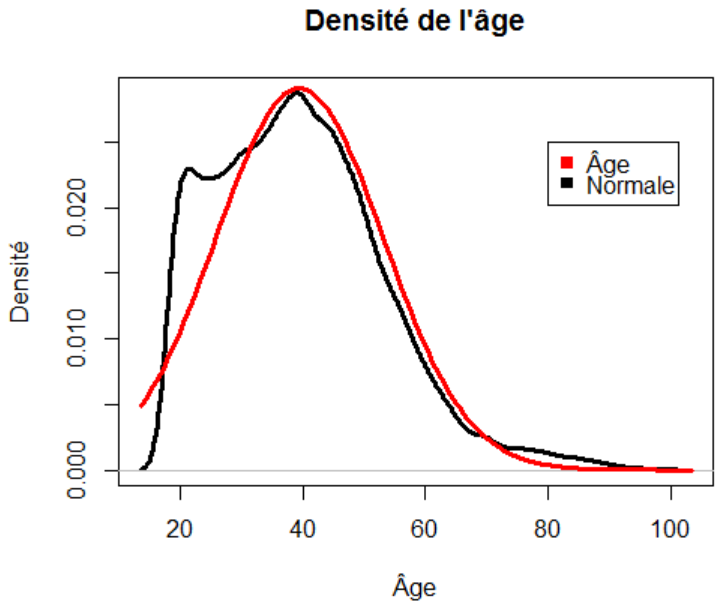


FIGURE 4.4 – Densité de l'âge et gaussienne associée

À partir de 30 ans, la courbe est suffisamment proche d'une loi normale. Si nous considérons nos données gaussiennes et que nous procédons à un test t de student, avec pour hypothèse nulle le fait que la moyenne de nos données soit celle de l'ensemble de la Wallonie, nous obtenons une p-valeur inférieure à 10^{-16}. Par conséquent, nous pouvons bien affirmer que les deux moyennes ne sont pas les mêmes et que les patients venant pour la première fois dans un SSM sont en moyenne plus jeunes que la population wallonne. Rappelons que cela ne signifie pas que les patients des SSM sont jeunes, puisque nous n'avons pas d'informations sur les patients qui y retournent.

Passons à la nationalité. Rappelons que les services de santé mentale présents dans la base de données sont ceux qui sont subventionnés par la région wallonne. Ils se trouvent donc tous en Wallonie. À la FIGURE 4.5, nous pouvons voir les différentes nationalités

des patients. Pour les histogrammes à venir, nous avons considéré uniquement les patients dont on avait l'information, donc sans les "DM". Nous avons ajouté le nombre de données considérées sur chaque graphique (c'est le "N"). L'analyse réalisée dans le chapitre précédent nous montre que les données manquantes ne posent pas de problème dans notre analyse. Cette FIGURE 4.5 nous permet de voir qu'une grosse majorité, soit 88.2%, des nouveaux patients sont de nationalité belge. De plus, parmi les non-belges il y a légèrement plus de personnes de nationalité hors union européenne qu'européenne.

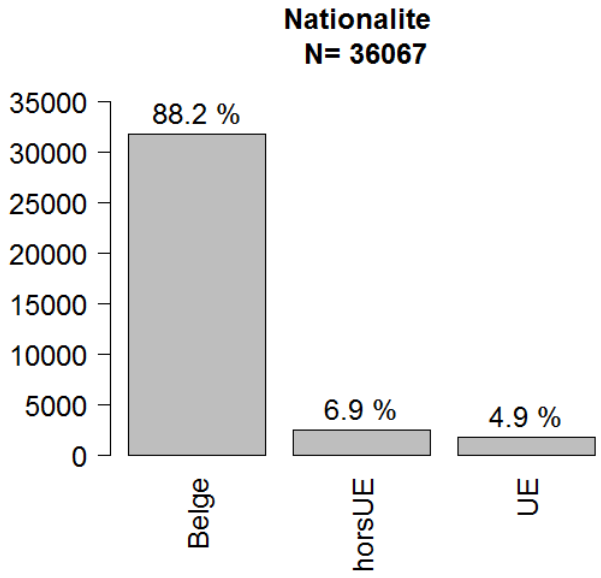


FIGURE 4.5 – Histogramme des nationalités

Désormais, sachant les proportions des différentes nationalités, nous pouvons nous interroger sur les langues maternelles de ces patients. À la FIGURE 4.6, nous pouvons voir que la plupart des patients, soit plus de 80% d'entre-eux, ont comme langue maternelle le français. Ensuite, nous avons des personnes dont la langue maternelle est une autre que celles proposées. Nous n'avons pas fait de graphique des autres langues, car aucune d'entre-elles ne revenait suffisamment de fois. Parmi les langues retenues, les deux autres langues nationales sont les moins présentes avec l'anglais. En effet, plus de patients parlent chez eux l'arabe, le turc ou l'italien que l'allemand, le néerlandais ou l'anglais.

En analysant cette variable, nous nous sommes rendu compte qu'il serait plus pertinent de savoir également si la personne parle le français. En effet, même si la langue maternelle n'est pas le français, il est possible que le patient sache s'exprimer dans cette langue. Si, par contre, ce n'est pas le cas, cette information est pertinente car cela signifie que le patient a eu recours à un traducteur pendant la consultation. Généralement, ce type de séances est pratiqué dans les services "Exil".

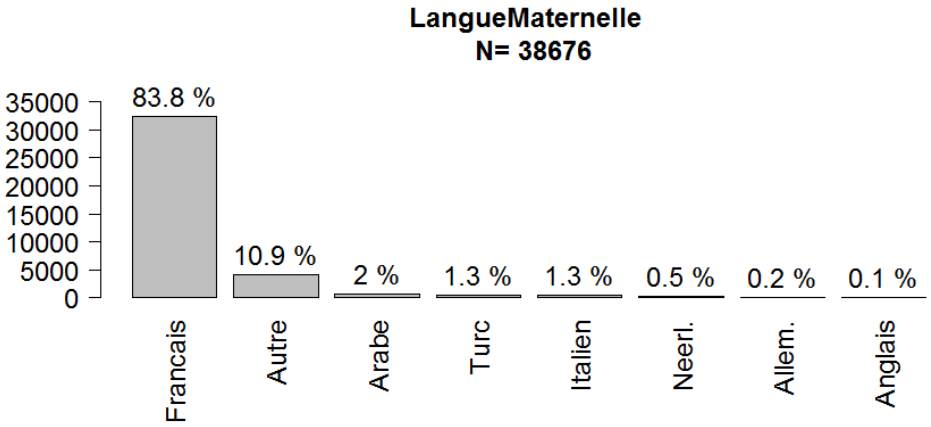


FIGURE 4.6 – Histogramme des langues maternelles

Continuons cette description avec l'état civil. À la FIGURE 4.7, nous pouvons voir que 44.3% des nouveaux patients sont enregistrés comme "célibataire". Ensuite, 28.1% sont mariés et 13.8% divorcés. Nous avons alors une minorité de personnes séparées, veuves ou ayant fait un contrat de vie commune.

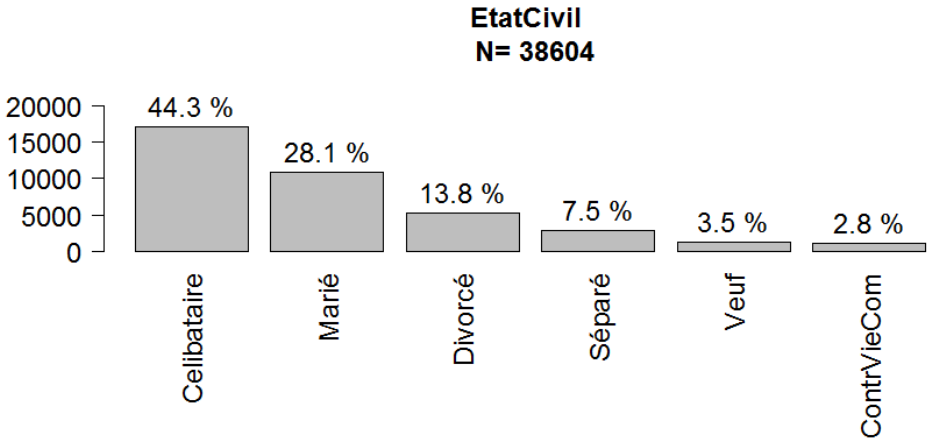


FIGURE 4.7 – Histogramme des états civils

Cette donnée est intéressante, mais ne nous donne pas d'idée du cadre réel dans lequel le patient vit. Il peut être célibataire aux yeux de la loi, mais vivre avec un compagnon/une compagne. Cette information est disponible dans la variable "Mode de vie", que nous avons représentée à la FIGURE 4.8. Nous pouvons voir que 60.6% des nouveaux patients, c'est-à-dire la majorité, vivent dans un milieu familial. Néanmoins, plus d'un quart vivent seuls. Enfin, les autres types de mode de vie sont minoritaires. Remarquons que cela n'est pas surprenant, car ces modes de vie semblent peu courants dans la Wallonie entière et pas uniquement au sein des patients des SSM.

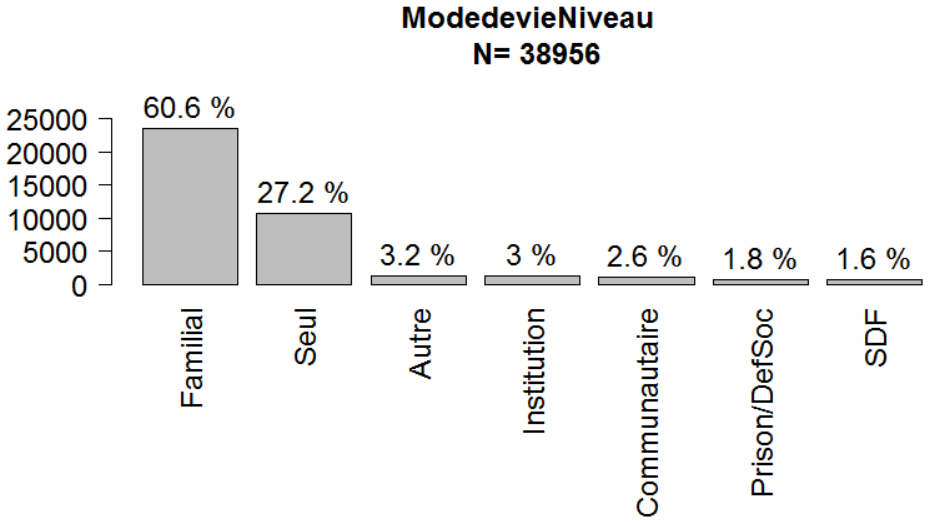


FIGURE 4.8 – Histogramme des modes de vie

Puisque la majorité vit dans un milieu familial, on peut alors se demander quelle est la configuration de ce milieu familial : avec des enfants, un partenaire, les parents,... ? La FIGURE 4.9 répond à cette question. En effet, nous avons repris uniquement les personnes vivant dans un milieu familial et tracé l'histogramme des modes de vie exacts.

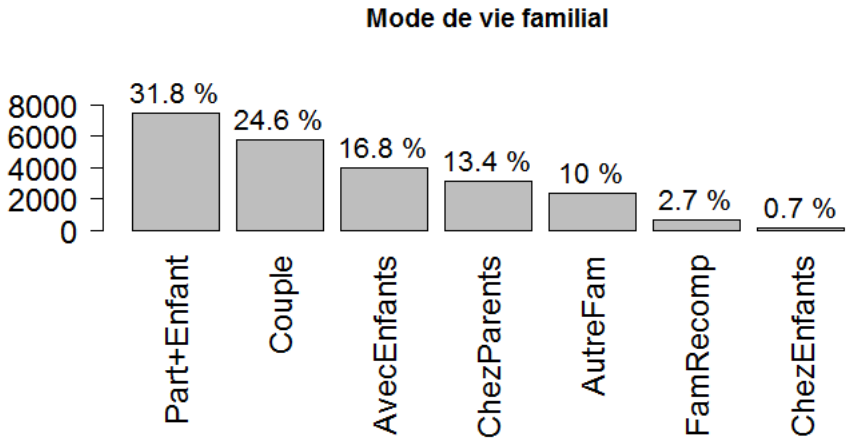


FIGURE 4.9 – Histogramme des modes de vie de type "Familial"

Nous pouvons observer que 31.8% d'entre-eux vivent avec un partenaire et des enfants. Un peu moins d'un quart vivent uniquement en couple, sans enfants, alors que 16.8% cohabitent avec des enfants, mais sans partenaire. Les personnes vivant chez leur parents représentent 13.4%. Enfin, nous avons 2.7% de familles recomposées et 0.7% qui sont chez leurs enfants. Remarquons que pour 10% des personnes vivant dans un milieu familial, nous n'avons pas d'informations complémentaires.

Passons aux variables nous donnant des informations sur le milieu professionnel dans lequel les nouveaux patients vivent. À la FIGURE 4.10, nous pouvons voir le niveau de scolarité atteint. Cette variable est formée de sorte que les personnes dont la dernière année réussie est, par exemple, la troisième secondaire soit répertoriées dans "secondaire". Par conséquent, "secondaire" ne signifie pas que le patient a obtenu son CESS, mais simplement qu'il a commencé ses secondaires.

Les patients n'étant pas arrivés jusqu'en secondaire représentent 8.8% de ceux dont on connait cette donnée. Nous avons 61.5% qui ont au moins commencé des études secondaires, donc plus de la majorité. Nous pouvons également voir que 26.7% ont entamé des études supérieures (réussies ou non), et 1.8% ont fait un apprentissage.

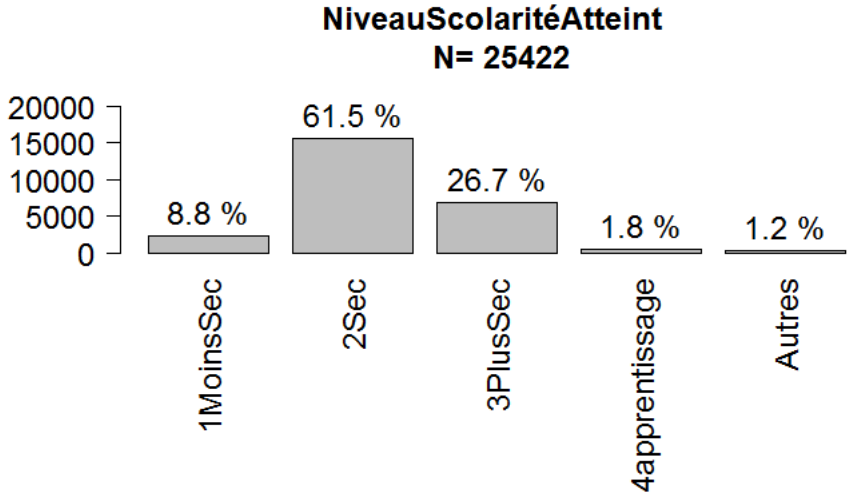


FIGURE 4.10 – Histogramme du niveau de scolarité atteint

La plupart des patients ont au moins commencé des études secondaires sans se lancer par la suite dans des études supérieures. Regardons jusqu'où ils sont allés dans ces études. Nous avons pour cela extrait les personnes dont la variable représentée ci-dessus était "secondaire" et affiché la dernière année réussie. Remarquons que pour certaines études secondaires techniques, une septième (voire une huitième) année est possible. De plus, la réponse "9" signifie que la personne a un certificat de qualification de l'enseignement spécial. Nous illustrons cette donnée, ainsi que le type de secondaire, à la FIGURE 4.11.

Nous constatons que 43.7% des patients ayant atteint comme plus haut niveau les secondaires ont pour dernière année réussie leur sixième, et 15.4% leur troisième. Pour ce qui est du type de secondaire, nous pouvons voir que 33% de ces patients suivaient des secondaires professionnelles.

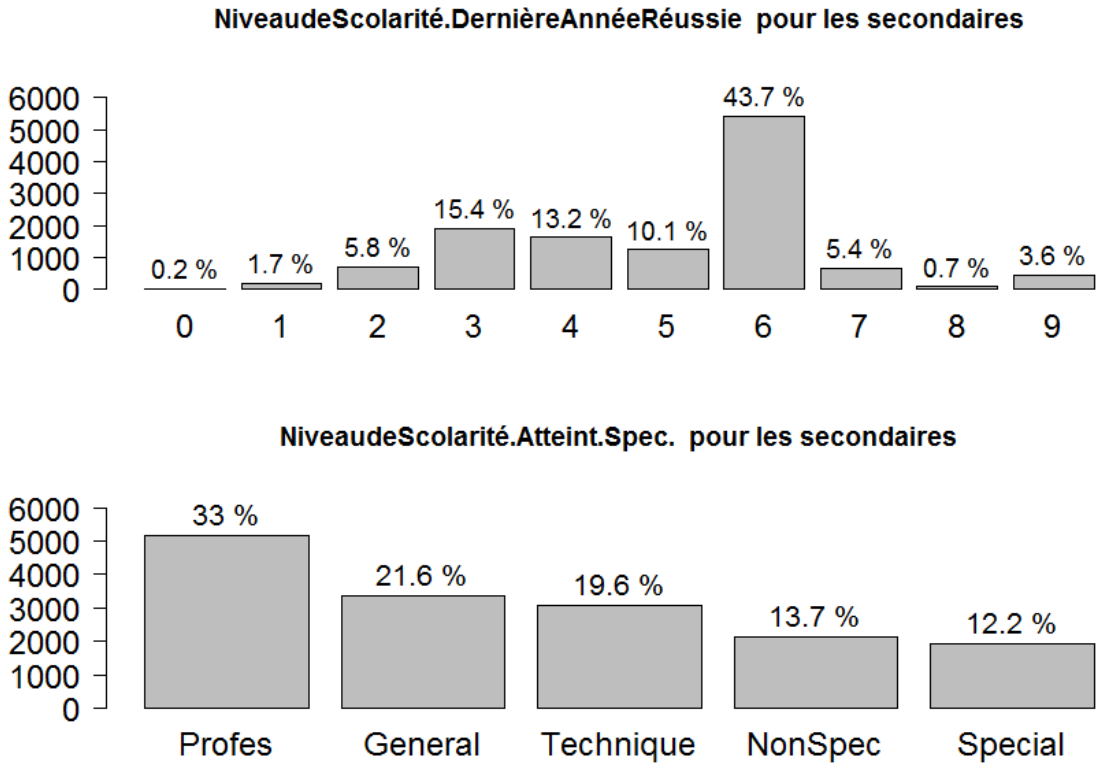


FIGURE 4.11 – Histogramme des types de secondaire et dernière année réussie

Le niveau d'étude est une variable intéressante, mais qui nous donne uniquement une information sur les études faites par le patient et non sur le métier qu'il exerce pour le moment. Cette information est disponible à la variable "Catégorie professionnelle". Pour bien comprendre cette variable, voici un extrait du manuel épidémiologique créé afin que les services remplissent correctement le formulaire : *Il s'agit d'indiquer ici à quelle catégorie professionnelle principale appartient le consultant (ou appartenait si la personne est actuellement inactive : pensionnée, malade, licenciée...) et ceci indépendamment du fait qu'il pratique actuellement son activité professionnelle ou pas.* [12]

À la FIGURE 4.12, nous pouvons voir que la catégorie professionnelle la plus présente, comprenant 35.5% des patients, est "sans profession". De plus, un quart des personnes dont on connait cette donnée sont des employés et 23.5% des ouvriers. Nous avons ensuite 8% d'étudiants, donc qui n'ont pas encore de profession mais sont déjà adultes ; 3.7% d'indépendants et très peu de cadres et directeurs ou de personnes de profession libérale.

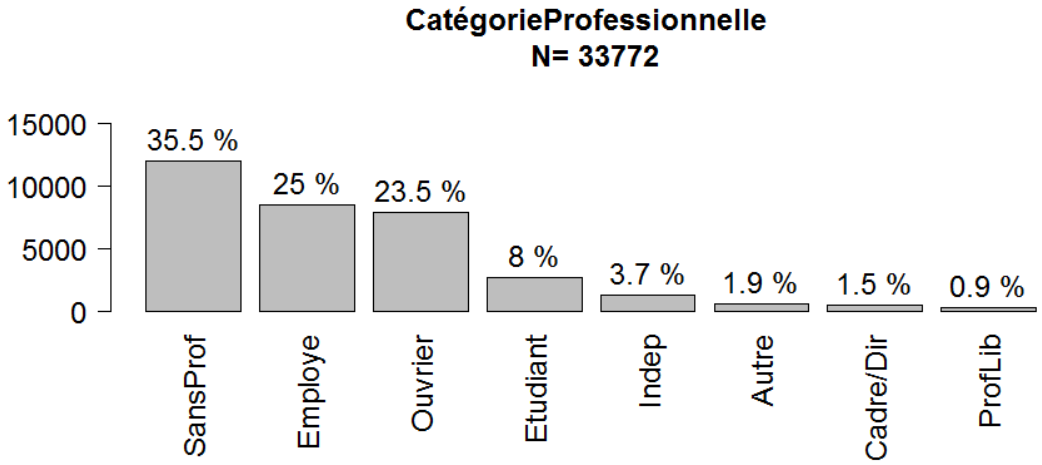


FIGURE 4.12 – Histogramme des catégories professionnelles

La variable suivante du profil des patients est la principale source de revenu, que l'on peut observer à la FIGURE 4.13. En addition aux études et à la profession, cette donnée nous permet de voir comment la personne paye ses factures, si elle travaille pour le moment ou pas, et si elle a accès à des aides sociales pour s'en sortir.

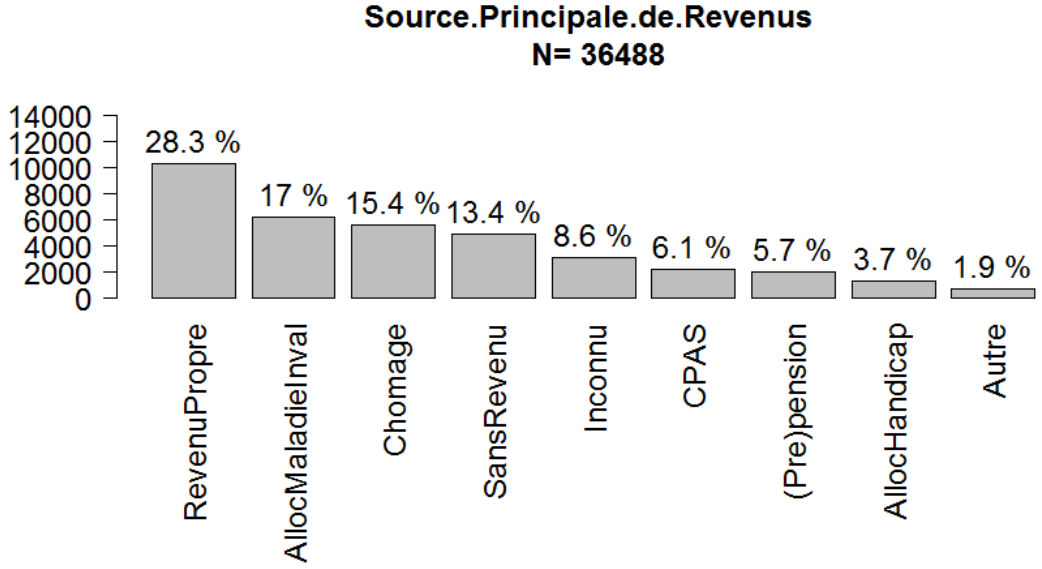


FIGURE 4.13 – Histogramme des sources principales de revenu

Soulignons le fait que c'est la source de revenu **principale**, donc si la personne peut entrer dans deux catégories, elle n'est inclue que dans celle qui lui rapporte le plus. Nous pouvons voir sur la FIGURE 4.13 que seuls 28.3% des patients ont pour rentrée

principale leur propre revenu. Malgré tout, cette source de revenu est la catégorie la plus représentée. La deuxième source principale de revenu est, et ce pour 17% des encodés, l'allocation de maladie et d'invalidité. Ensuite, 15.4% bénéficient d'allocations de chômage, et 13.4% sont sans revenu. Cela peut sembler étonnant puisque nous avions seulement 8% d'étudiants. Il est alors normal de se demander comment vivent ces personnes. Nous analyserons donc leur mode de vie par la suite. Nous avons enfin le CPAS (6.1%), la pension (5.7%)et les allocations "handicapés" (3.7%).

Afin d'observer plus précisément le cadre de vie des personnes sans revenu, nous avons tracé leur mode de vie et leur catégorie professionnelle. Commençons par regarder leur mode de vie à la FIGURE 4.14. Nous pouvons voir que la majorité de ces patients vivent dans un milieu familial. Ensuite, nous avons les personnes en prison ou en défense sociale qui représentent 8.9%. Nous avons alors 8.8% d'entre-eux qui vivent seuls et ensuite les patients vivant dans un milieu communautaire, les "autres", en institution, les sans domicile fixe et enfin, 1% de données manquantes.

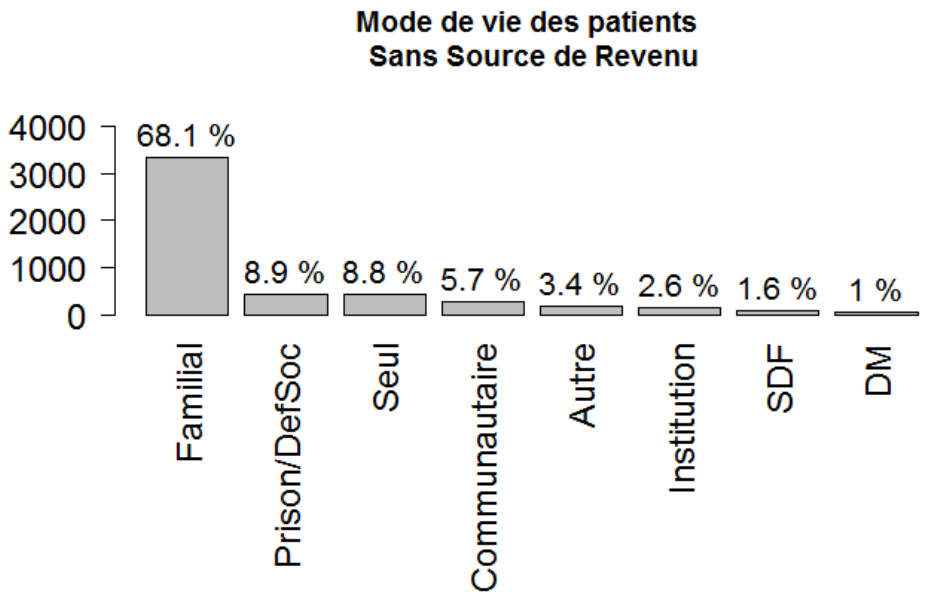


FIGURE 4.14 – Histogramme des modes de vie des patients sans revenus

Passons aux catégories professionnelles des personnes sans revenu. Nous pouvons voir l'histogramme représentant cette variable à la FIGURE 4.15. Nous pouvons constater que la plupart de ces patients sont soit sans profession, soit étudiant. Les autres catégories professionnelles ne sont que peu représentées par rapport à ces deux-ci. Remarquons qu'il est tout à fait logique que les étudiants n'aient pas encore de revenu.

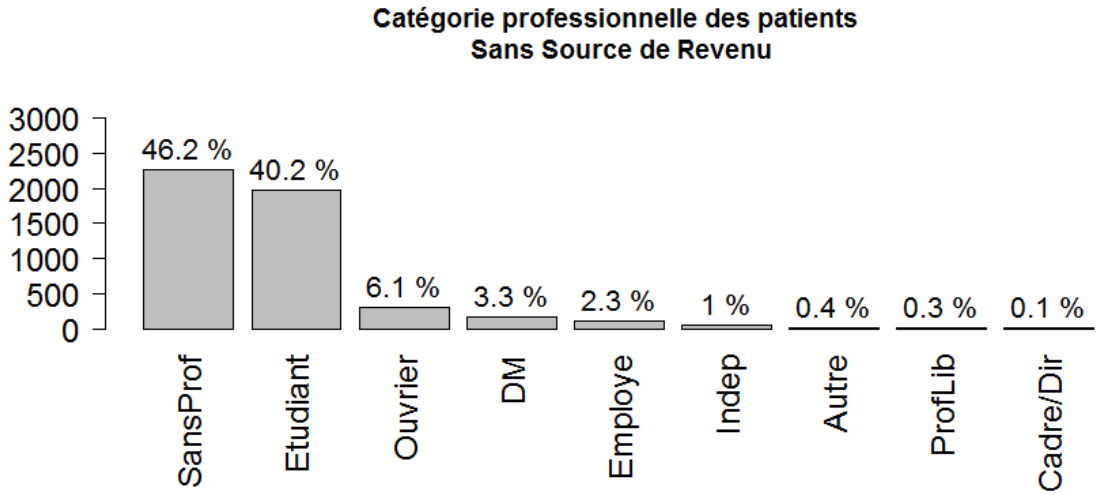


FIGURE 4.15 – Histogramme des catégories professionnelles des patients sans revenu

La dernière donnée socio-démographique que nous avons sur les nouveaux patients est la province dans laquelle les patients habitent. Nous pouvons voir l'histogramme correspondant à la FIGURE 4.16. Nous constatons que les provinces wallonnes sont plus présentes, ce qui est compréhensible puisque tous les SSM qui ont rempli les questionnaires sont en Wallonie.

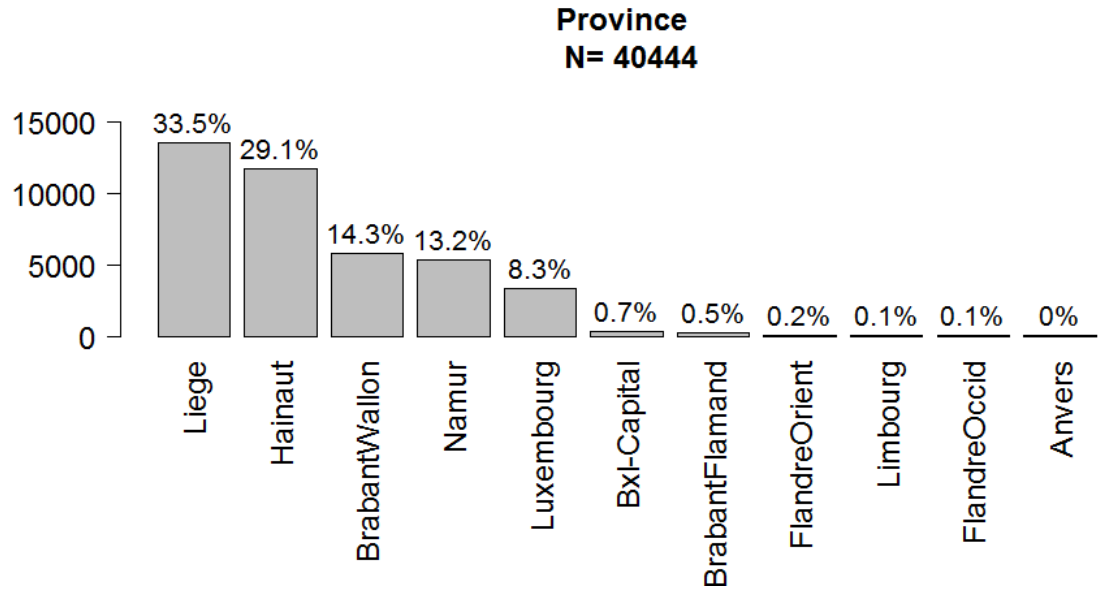


FIGURE 4.16 – Histogramme des provinces

Liège et le Hainaut sont les provinces les plus représentées. Malgré tout, il faudra rester prudent sur les conclusions. En effet, nous avons choisi de ne pas faire de carte avec les communes précises, car divers facteurs entrent en considération et il n'est pas aisé de voir si certains lieux sont plus touchés. En effet, il y a le fait que les SSM sont plus ou moins éloignés des logements des patients. Si aucun service ne se trouve près du patient, il va peut-être hésiter à consulter, ou simplement renoncer. Par conséquent, il est possible que Liège soit plus représenté non pas parce que les habitants ont plus vite des problèmes, mais parce qu'il y a plus de services disponibles et que cela incite les personnes à s'y rendre plus rapidement. Une étude plus poussée de cette variable devrait être faite pour tirer des conclusions géographiques, telle que l'attractivité des services, etc. Malheureusement, dans le cadre de ce mémoire, nous n'aurons pas l'occasion de pousser cette question plus loin.

Chapitre 5

Profil médical

Le but de ce chapitre est de décrire le profil médical des patients. Nous séparerons celui-ci en deux grandes parties. Une première définissant ce que l'on appellera l'"itinéraire thérapeutique", qui reprendra les différentes étapes que le patient a déjà franchies tout au long de sa thérapie, telles que les prises en charge antérieures, mais également les caractéristiques de sa consultation, s'il est venu seul, ... La deuxième partie reprendra les "résultats" de la consultation, donc le diagnostic émis par le praticien et les prises en charge proposées.

5.1 Itinéraire thérapeutique

Passons maintenant aux variables concernant la (ou les) consultation(s), ainsi que la raison de la venue du patient. Commençons par analyser les types de dossiers, c'est-à-dire si le dossier est individuel, de couple ou familial. À la FIGURE 5.1, nous pouvons voir la répartition de cette donnée. Cette illustration nous permet de voir que 93.8% des dossiers sont individuels. Seuls 3.9 et 2.2% sont respectivement de type "couple" et "familial".

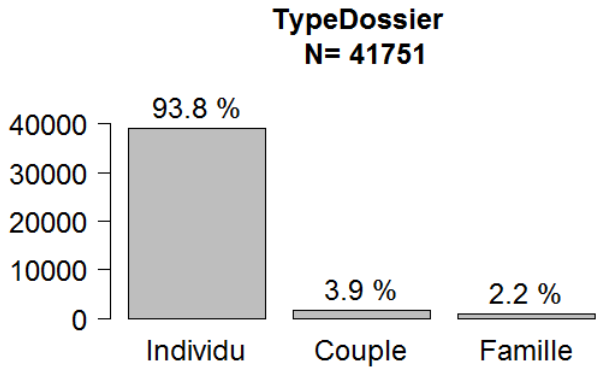


FIGURE 5.1 – Histogramme des types de dossier

Une question pertinente dans cette étude épidémiologique est la nature et l'origine de la démarche des patients. Cette première variable définit si la personne a décidé seule de venir ou si quelqu'un l'a aidée ou obligée à prendre cette décision. À la FIGURE 5.2, nous pouvons voir la nature des démarches des patients. Parmi les données récoltées, 64.9% sont des démarches orientées ; 28.2% sont spontanées et 6.9% contraintes.

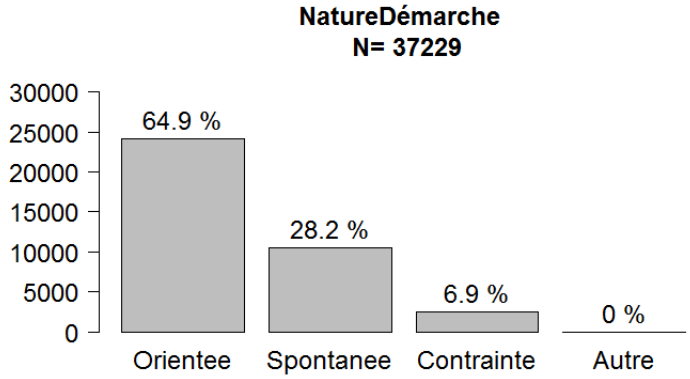


FIGURE 5.2 – Histogramme de la nature des démarches

Il faut rester prudent sur les conclusions de cette variable, car sa définition était un peu ambigüe. En effet, une personne qui n'est pas bien et en parle à son ami qui lui conseille de consulter peut se trouver dans les spontanées ou les orientées en fonction de la définition de la spontanéité propre à chaque personne. Pour savoir qui a pu encourager le patient à se rendre dans un SSM, nous avons l'origine de la démarche à la FIGURE 5.3.

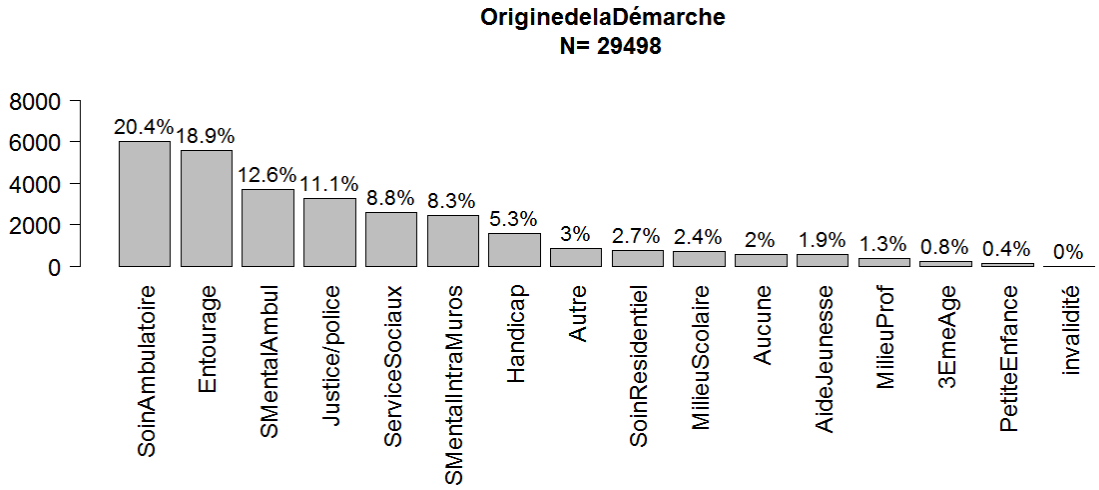


FIGURE 5.3 – Histogramme de l'origine de la démarche

Pour bien comprendre cette variable, nous avons copié la ligne lui correspondant dans le manuel d'utilisation des données épidémiologiques : *Il s'agit en fait de mentionner ceux que l'on désigne communément comme les "envoyeurs".* [12] Pour mieux voir ce que signifie chaque dénomination, il y a en annexe un tableau reprenant les sous-branches comprises dans chaque catégorie. Nous observons 20.4% de nouveaux patients envoyés par des soins ambulatoires. L'entourage a envoyé 18.9% des personnes. De plus, 12.6% ont comme origine de la démarche un soin de santé mentale ambulatoire. Par conséquent, si on regroupe les deux types de santé ambulatoire, nous avons 33% des patients. Ensuite, 11.1% arrivent là incités par la justice ou la police ; 8.8% par les services sociaux et 8.3% par des services de santé mentale intra-muros. Les autres expéditeurs ne sont que peu représentés.

Nous nous sommes alors intéressés aux combinaisons de nature et d'origine de la démarche les plus importantes. Pour cela, nous avons, à la TABLE 5.1, le tableau de contingence de ces variables, avec, en rouge, les plus gros effectifs dans l'origine, pour chaque nature. Remarquons que nous n'avons pas laissé les "DM" dans ce tableau. Cependant, la majorité des données manquantes dans l'origine de la démarche font partie des démarches spontanées. Intuitivement, nous nous attendions à cela, puisque personne ne les envoie. De plus, nous avons beaucoup de cas où la donnée manque dans les deux variables.

Origine\Nature	Contrainte	Orientée	Spontanée
3EmeAge	1	154	13
AideJeunesse	54	452	41
Aucune	2	20	480
Autre	52	608	192
Entourage	33	2948	2456
Handicap	43	1420	70
Justice/police	1973	1117	137
MilieuProf	13	331	45
MilieuScolaire	3	632	74
PetiteEnfance	4	110	11
ServicesSociaux	28	2212	313
SMentalAmbul	31	3127	490
SMentalIntraMuros	29	2163	207
SoinAmbulatoire	21	5369	507
SoinResidentiel	4	726	46

TABLE 5.1 – Table de contingence de la nature et l'origine de la démarche

Nous constatons que la plupart des personnes contraintes l'ont été par la justice ou la police. De plus, la plupart des personnes dont la démarche était spontanée ont pour origine l'entourage. Enfin, parmi ceux dont la démarche avait été orientée, la majorité ont été orientés par un soin ambulatoire, et beaucoup l'ont également été par la santé mentale ambulatoire (intra-muros et ambulatoire), les services sociaux et l'entourage.

Nous pouvons représenter la relation entre l'origine et la nature de la démarche grâce à un processus descriptif nommé "Analyse Factorielle des Correspondances". Cette méthode à l'avantage de ne pas nécessiter de modèle de causalité. On n'essaye pas d'expliquer une donnée par une autre, mais juste de décrire les deux variables simultanément et le plus clairement possible. L'explication qui va suivre sera basée sur la source [25].

Globalement, cette analyse consiste à placer les réponses à ces deux questions dans un hypercube, pour ensuite trouver la meilleure "vue". En effet, nous ne représenterons cet hypercube que dans deux dimensions, donc nous allons certainement perdre de l'information. Chaque coin de cet hypercube correspond à une combinaison de nature et d'origine de la démarche. Puisque certaines combinaisons sont plus représentées que d'autres, certains coins auront plus de poids.

Nous appelons "inertie totale" une mesure de l'écart entre les données récoltées et ce que l'on aurait dû observer si les variables avaient été indépendantes. Le processus va essayer de conserver un maximum d'inertie expliquée, pour que l'on perde le moins d'informations possible en remettant tout ça en 2 dimensions. Chaque axe représente donc une proportion de l'inertie expliquée.

Pour faire cela, nous utiliserons dans le logiciel "R" une fonction appelée "Multiple Correspondance Analysis" et notée "MCA". Remarquons que c'est une "**Multiple** Correspondance Analysis", car celle-ci est applicable à plus de deux variables qualitatives.

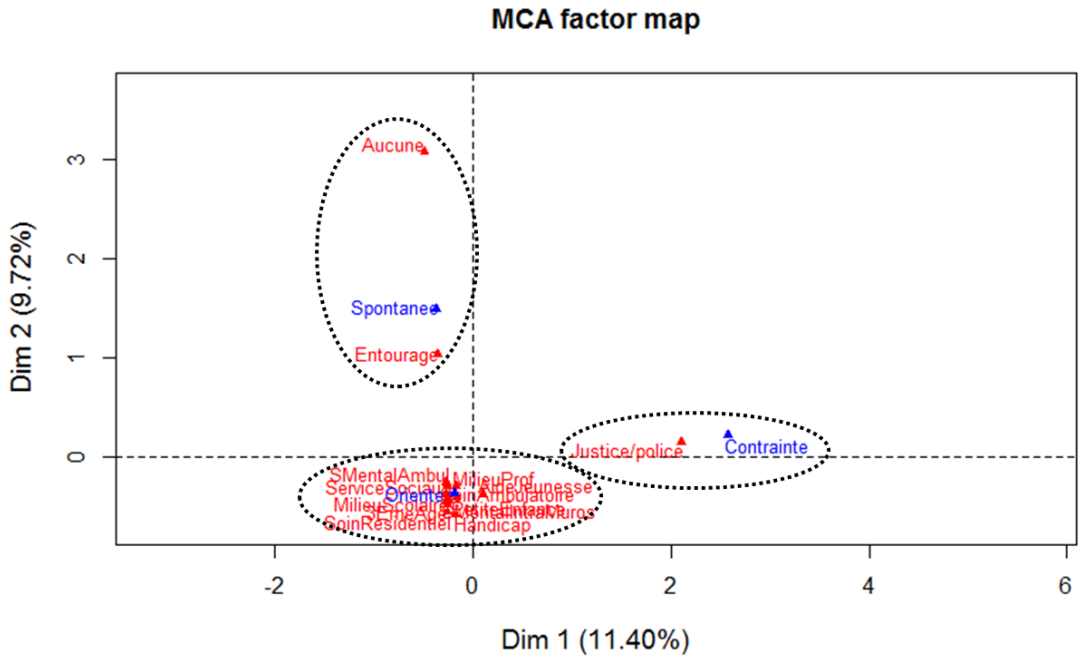


FIGURE 5.4 – Analyse en correspondances multiples : l'origine et la nature

La FIGURE 5.4 comprend cette MCA pour l'origine et la nature de la démarche. Nous pouvons voir que nous avons 21.12% de la variance expliquée au total de ces deux axes. De plus, la matrice possédait 15 valeurs propres, donc pour expliquer 100% de l'inertie totale, il aurait fallu au moins 15 dimensions. Sur notre FIGURE 5.4, nous voyons que le premier axe, qui explique le plus la variance, est principalement influencé par la nature de la démarche "contrainte" qui se détache vraiment des autres. L'angle créé entre la modalité "contrainte" et la dimension 1 est petit, donc son cosinus au carré a une valeur proche de 1. Cette mesure du carré des cosinus permet d'identifier les points les plus proches des axes. Nous avons observé les dimensions supérieures à deux, et celles-ci n'étaient pas aussi fort guidées par une catégorie. En effet, tous les cosinus au carré des angles créés avec l'axe étaient assez petits.

Maintenant que toutes ces notions sont définies, nous pouvons interpréter le graphique. Nous pouvons voir sur la FIGURE 5.4 que les données semblent pouvoir être divisées en trois groupes. Afin d'illustrer clairement les trois classes, nous les avons entourées. Nous voyons que les 11.4% de variance expliqués par la première dimension concernent les démarches contraintes par la police ou la justice. Ensuite, la deuxième dimension expliquant 9.72% de l'inertie reprend les personnes qui ont eu une démarche spontanée guidée par l'entourage, voir non guidée. Enfin, les autres modalités forment un noyau opposé aux deux groupes déjà définis et proches de l'origine. Ces modalités sont différenciées dans les dimensions supérieures, mais nous avons déjà plus de 20% de variance expliquée ici. Ce noyau reprend les démarches orientées, peu importe le service qui les a encouragé à se rendre dans un SSM.

Maintenant que nous avons observé la démarche suivie par le patient, voyons la cause de sa venue, ses motifs. Dans un premier temps, nous regarderons la "famille" de ce motif, donc son caractère personnel ou relationnel. À la FIGURE 5.5, nous pouvons voir que 71.5% des patients viennent pour des problèmes personnels et 25% pour des problèmes relationnels. Seuls 2.6% ont d'autres motifs et 0.9% n'en ont aucun.

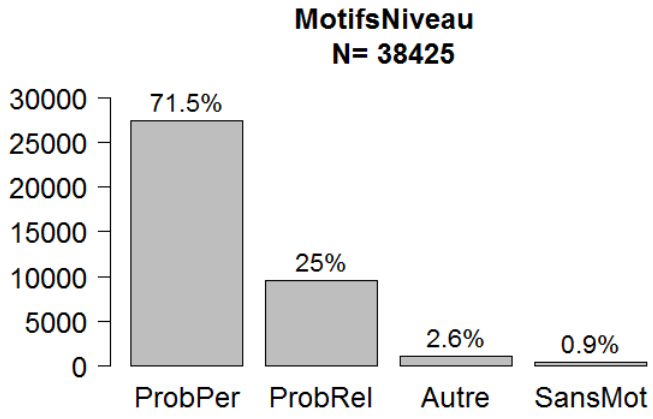


FIGURE 5.5 – Histogramme des motifs de la première consultation (niveaux)

Nous avons des informations supplémentaires pour chaque patient. En effet, nous avons un motif spécifique nous montrant plus précisément ce qui a poussé le patient à venir dans un SSM. Chaque motif spécifique entre soit dans la catégorie "personnelle", soit dans la catégorie "familiale". Afin de mieux établir les motifs, regardons d'abord la proportion de chaque motif spécifique au sein de chaque catégorie générique, et ensuite, rassemblons tous les types spécifiques sur une image. Nous trouvons cela à la FIGURE 5.6.

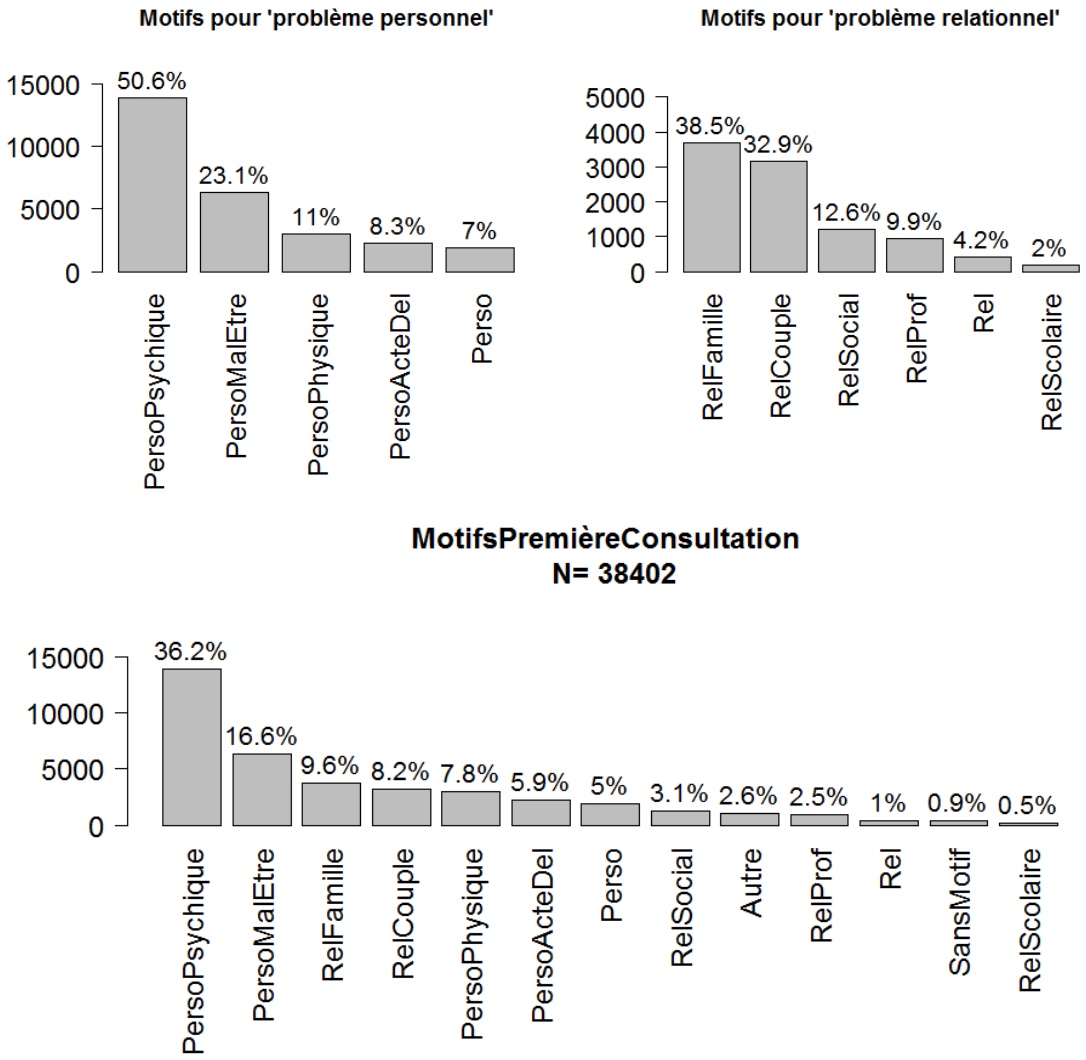


FIGURE 5.6 – Histogramme des motifs de la première consultation (niveaux généraux)

Nous pouvons voir que, parmi les problèmes personnels, 50.6% d'entre-eux sont psychiques et 23.1% proviennent d'un mal-être. En considérant toutes les sortes de problèmes, ces deux-ci restent les principaux, concernant respectivement 36.2% et 16.6% de l'ensemble des patients. Après ces deux motifs, l'ensemble des patients passent aux problèmes relationnels familiaux englobant 38.5% des problèmes relationnels et 9.6% de tous les problèmes. Viennent alors les problèmes de couple qui concernent 8.2% de tous

les motifs, et 32.9% des motifs relationnels. Les suivants, les problèmes physiques représentent 11% des personnels et 7.8% de tout. Remarquons que, pour 7% des problèmes personnels et 4.2% des relationnels, nous n'avons pas d'informations complémentaires. Enfin, les problèmes relationnels sociaux, professionnels ou scolaires ne sont que peu présents.

Nous savons désormais ce qui pousse les patients à se rendre dans un SSM. Il est temps de nous intéresser à ce qu'ils attendent de cette consultation, leur espoir ou leur demande. À la FIGURE 5.7, nous pouvons voir les demandes des consultants.

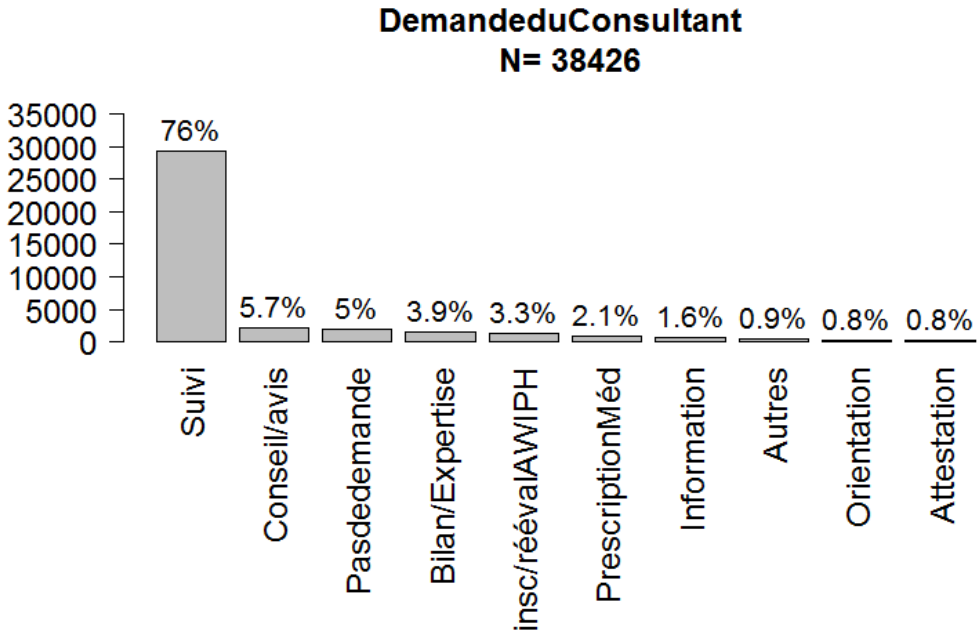


FIGURE 5.7 – Histogramme des demandes des consultants

Analysons cette illustration. Nous avons 76%, soit plus de trois quarts, des patients qui désirent un suivi. De plus, 5.7% sont venus pour avoir un conseil ou un avis et 5% n'avaient aucune demande. Nous pouvons également voir que seulement 2.1% des nouveaux patients se sont rendus dans un SSM dans le but d'obtenir une prescription médicale. C'est plutôt rassurant, car ça signifie que peu de patients s'y rendent pour avoir une prescription.

Au niveau des prises en charge, observons celles qui sont déjà intervenues avant la première venue dans un SSM. À la FIGURE 5.8, nous pouvons voir l'histogramme correspondant à cette donnée.

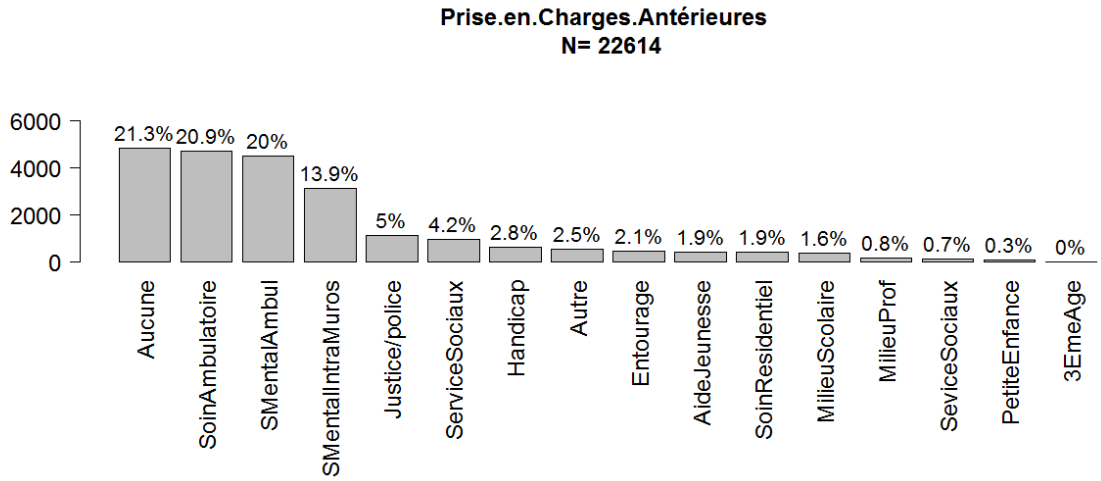


FIGURE 5.8 – Histogramme des prises en charge antérieures

Nous constatons que 21.3% des nouveaux consultants n'étaient pas du tout pris en charge avant. Ensuite, 20.9% avaient déjà été aidés par des soins ambulatoires et 20% par des soins de santé mentale ambulatoires. Cela nous permet de voir que les soins de type ambulatoire que ce soit généraux ou de santé mentale sont les plus présents au sein des prises en charge antérieures. Nous avons alors les soins de santé mentale intra-muros avec 13.9%. Les autres catégories ne sont que très peu présentes.

Nous pouvons passer à la donnée suivante. Chacun a généralement des ressources extérieures à ce milieu professionnel, telles que son entourage, sa famille ou autre. Nous avons également cette donnée. Remarquons que cette variable permettait plusieurs réponses (maximum 3). Elles ont alors été encodées par ordre d'importance. Nous pouvons, dans un premier temps, regarder uniquement la proportion de personnes ayant plusieurs ressources. Nous avons les ressources principales de 30275 nouveaux patients et parmi eux, 17577 ont une deuxième ressource. Il y a donc 12698 personnes n'ayant qu'une seule ressource. De la même manière, 11844 nouveaux patients ont exactement 2 ressources extérieures et 5733 en ont 3. Nous pouvons calculer la proportion de patients ayant le même nombre de ressources à la TABLE 5.2.

Une ressource	deux ressources	trois ressources
42%	39%	19%

TABLE 5.2 – Table des nombres de ressources

Remarquons qu'il n'y avait pas la possibilité de remplir "aucune" pour cette variable. Par conséquent, notre étude se base ici uniquement sur les patients qui avaient au moins l'accès à une ressource extérieure. Parmi eux, 42% n'en possédait qu'une et 39% deux. Seulement 19% des personnes encodées pouvaient se tourner vers trois ressources extérieures différentes. Regardons désormais la FIGURE 5.9 afin de déterminer la nature de chaque ressource. Les ressources principales sont majoritairement personnelles, les

deuxièmes ressources extérieures sont surtout familiales et en troisième position, nous avons plus d'amis. Nous voyons donc ici qu'il est intéressant d'avoir trié les ressources par ordre d'importance.

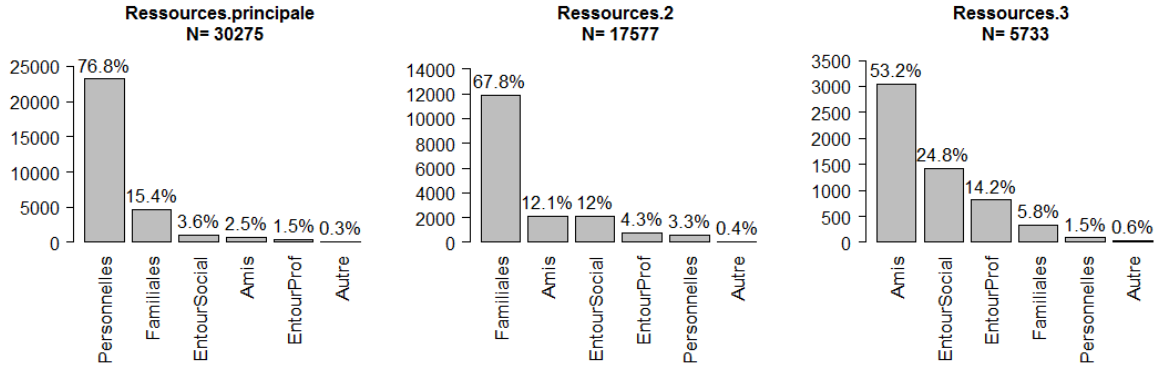


FIGURE 5.9 – Histogramme des ressources extérieures

Maintenant que nous avons observé séparément chaque ressource, il est intéressant de joindre ces données afin de voir le pourcentage de patients ayant accès à chaque ressource. A la FIGURE 5.10, nous pouvons voir l'ensemble des ressources extérieures avec la proportion associée. Par conséquent, nous pouvons interpréter la première colonne de la façon suivante : parmi tous les patients dont nous connaissons au moins une ressource, 79% ont des ressources personnelles. Ensuite, les ressources familiales sont accessibles pour 55.8% des patients. Donc, il faut rester prudent avec cet histogramme car chaque colonne s'interprète individuellement. Par construction, il est normal que la somme de toutes les colonnes fasse plus que 100% étant donné que certains patients ont accès à plusieurs types de ressources. Nous constatons qu'au total, la ressource la plus présente est la ressource personnelle, suivie de la ressource familiale. Viennent alors les amis avec seulement 19.6%.

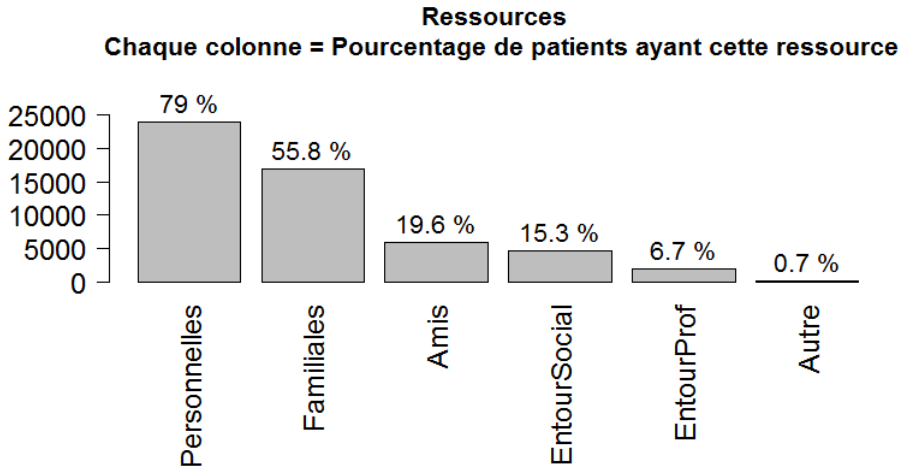


FIGURE 5.10 – Histogramme de l'ensemble des ressources extérieures

En plus des ressources extérieures des patients, nous avons les réseaux professionnels. Au niveau de cette variable, nous pouvons lire dans le manuel épidémiologique : "*L'objectif de cette variable est de connaître les autres professionnels qui sont également impliqués dans une intervention avec le consultant en même temps que le SSM. La question à laquelle tente de répondre cette variable est de savoir quelles sont les ressources professionnelles dont dispose le consultant.*"

Dans cette colonne, des erreurs d'encodage se sont produites pour les services de la province du Hainaut à partir du deuxième réseau professionnel. Il était possible d'intégrer 4 réseaux différents, mais seuls moins de 4% des patients en ont un troisième. De plus, de nombreuses erreurs d'encodage se trouvent dans la deuxième colonne. J'ai donc décidé de me focaliser sur le réseau professionnel principal. La FIGURE 5.11 nous montre la répartition des différents réseaux professionnels principaux.

Cette donnée est inconnue pour 14.3% des nouveaux patients. Remarquons que tous les services n'ont pas utilisé la possibilité de "Inconnu" et ont soit inséré un réseau professionnel, soit laissé la donnée manquante. Il vaut donc mieux retirer les données inconnues au même titre que les manquantes. Nous pouvons voir que 36.2% ont déjà recours à des soins ambulatoires et 16.3% à des services de santé mentale ambulatoire. Ce qui nous fait un total de 52.5% ayant comme ressource des soins ambulatoires. Nous avons ensuite plus de 9% de personnes dans le réseau de la justice, ainsi que des services sociaux.

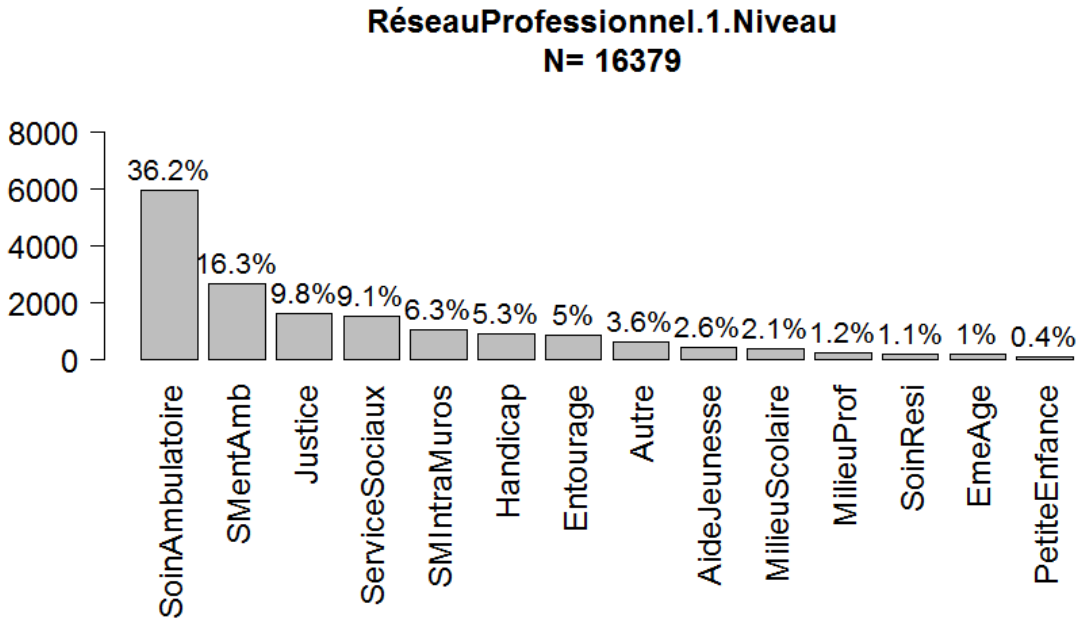


FIGURE 5.11 – Histogramme des réseaux professionnels

5.2 Problématiques rencontrées par le patient

Passons au diagnostic émis par le praticien. Pour ce faire, la base de données contient cette information sur base de codes "ICD10" qui permettent de classer les pathologies. Comme nous pouvons le voir dans le manuel épidémiologique, voici la classification des différents codes ICD10 (remise à jour avec l'aide du CRESAM) : [12](page18)

- *troubles mentaux et troubles du comportement (F00 à F99)*
- *diagnostic médical (de A00 à Y 99 à l'exclusion des codes F),*
- *et/ou diagnostic social (facteurs influant sur l'état de santé et motifs de recours aux services de santé)(Z00 à Z99).*

Cette classification nous montre déjà que la première lettre permet d'identifier la catégorie dans laquelle se trouve le code.

Cette question autorisait plusieurs réponses (jusqu'à 6), à nouveau triées par ordre d'importance. Nous avons décidé de ne garder que les 3 premiers diagnostics, car très peu de patients étaient concernés par plus de codes. Notons que les données manquantes ont ici été encodées avec des "0" et non des "DM" pour ne pas confondre avec les codes "D" lorsque nous ne considérerons que la première lettre.

Comme pour les ressources extérieures, il était possible de remplir une ou plusieurs colonnes pour cette variable en fonction du nombre de problèmes rencontrés par le patient. Il est donc, ici, intéressant d'observer la proportion des patients ayant exactement un, deux ou plus de codes.

Nous pouvons formuler cela sous forme de tableau. Nous constatons à la Table 5.3 ci-dessous que 47.4% des nouveaux patients n'ont qu'un seul code. Ensuite, 30.5% en ont exactement deux et 22.1% en ont trois ou plus.

Un code	deux codes	au moins trois codes
47.4%	30.5 %	22.1%

Table 5.3 – Répartition des nombres de codes

À la Figure 5.12, nous pouvons voir les premières lettres des trois codes ICD10 principaux. Nous pouvons voir que la majorité des codes principaux sont de type "F", alors que la majorité des codes mis en deuxième et troisième position sont de type "Z". Nous constatons également que les autres types de codes, "diagnostics médicaux" ne sont que minoritaires pour les 3 codes. Nous avons appelé ces codes "Autre" car ils contiennent toutes les lettres qui ne sont ni "Z" ni "F".

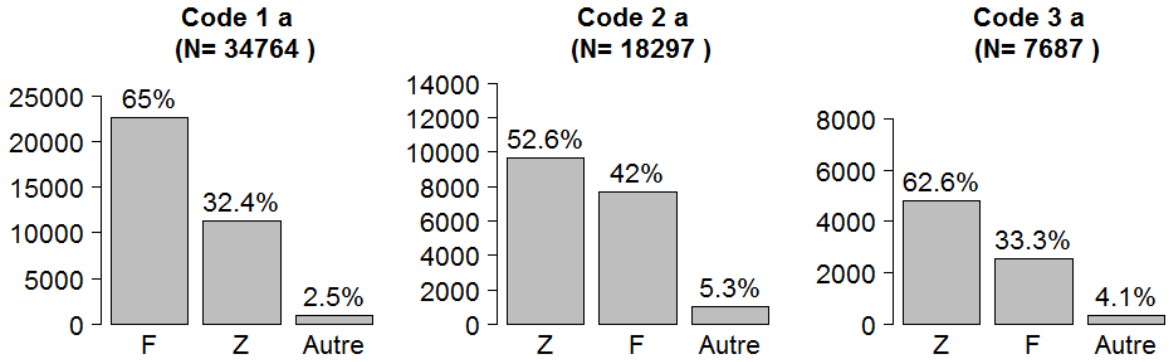


FIGURE 5.12 – Histogramme des lettres des codes ICD10

Si nous rassemblons ces 3 codes et que nous observons le pourcentage de patients, dont on a au moins 1 code, possédant chaque type de codes (lettres), nous obtenons la FIGURE 5.13. Il en ressort que 94.5% des patients possèdent un code F, que ce soit en première, deuxième ou troisième position. De la même manière, des codes Z sont présents au sein des trois premiers codes pour 74% des personnes encodées. Les autres codes sont à nouveau très peu présents.

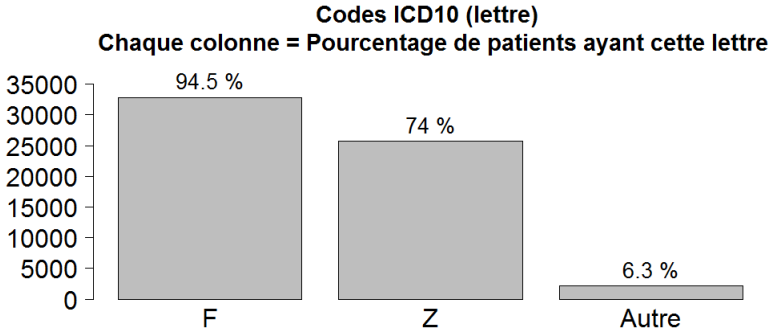


FIGURE 5.13 – Histogramme des lettres des codes ICD10 rassemblés

Si nous nous focalisons désormais sur les codes de type "F" et que nous regardons, au sein de tous les codes de ce genre (en 1ère, 2ème et 3ème place), nous pouvons voir la répartition de la FIGURE 5.14. Avec ce genre de codes, nous avons gardé uniquement les deux premières lettres car celles-ci donnent déjà une bonne idée du type de problème. En effet, nous pouvons voir dans le tableau de la TABLE 5.4 rédigé par le CRESAM pour ce mémoire, la classification des types "F" , ainsi que des exemples. Avant de ne regarder plus que les deux premiers caractères, nous avons changé les "F99", qui correspondaient à un trouble mentale sans précision, en "FA" pour "F Autre", afin de ne pas les confondre avec les autres "F9".

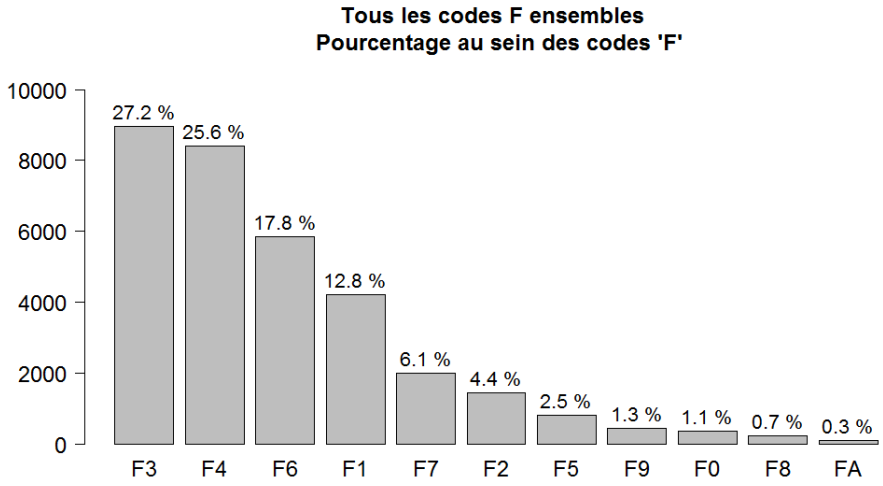


FIGURE 5.14 – Histogramme de tous les codes "F" rassemblés

	Intitulé	Exemples
F0	Troubles mentaux organiques, y compris les troubles symptomatiques	Maladie d'Alzheimer
F1	Troubles mentaux et troubles du comportement liés à l'utilisation de substances psychoactives	Troubles mentaux et troubles du comportement liés à l'utilisation d'alcool ou de cocaïne
F2	Schizophrénie, trouble schizotypique et troubles délirants	Schizophrénie, trouble psychotique
F3	Troubles de l'humeur (affectifs)	Episode dépressif, un trouble bipolaire
F4	Troubles névrotiques, troubles liés à des facteurs de stress, et troubles somatoformes	Phobies sociales, trouble obsessionnel-compulsif, réaction aigüe à un facteur de stress
F5	Syndromes comportementaux associés à des perturbations physiologiques et à des facteurs physiques	Anorexie mentale, terreurs nocturnes
F6	Troubles de la personnalité et du comportement chez l'adulte	Personnalité paranoïaque (à différencier d'une schizophrénie paranoïde), personnalité anxieuse
F7	Retard mental	
F8	Troubles du développement psychologique	Troubles du développement de la parole et du langage, trouble de la lecture, autisme infantile
F9	Troubles du comportement et troubles émotionnels apparaissant habituellement durant l'enfance ou à l'adolescence	Troubles hyperkinétiques, angoisse de séparation dans l'enfance,...

TABLE 5.4 – Exemples pour les codes F (troubles mentaux et troubles du comportement)

Nous pouvons voir que 27.2% des codes F sont des troubles de l'humeur et 25.6% des troubles névrotiques. Ensuite, nous avons 17.8% de troubles de la personnalité et du comportement. Ce sont les trois codes les plus présents au sein des codes F. Les codes F8, F9 et F0 ne sont que très peu présents ici. Cela peut s'expliquer par le fait que les codes F8 et F9 semblent plus destinés aux patients enfants ou adolescents. Pour le code F0, on voit simplement qu'il est moins présent.

Nous pouvons faire la même chose, mais cette fois en regardant à quelle "place" ce code était, afin de voir si certains codes reviennent plus dans le problème principal et d'autres dans les suivants. Nous observons à la FIGURE 5.15 que le code F le plus présent dans le premier code est le F3 ; alors que dans le deuxième code, le plus présent est le F6 ; et que dans le troisième, le plus présent est le F4. Les codes F3, F4 et F6 restent dans tous les cas les trois codes les plus présents.

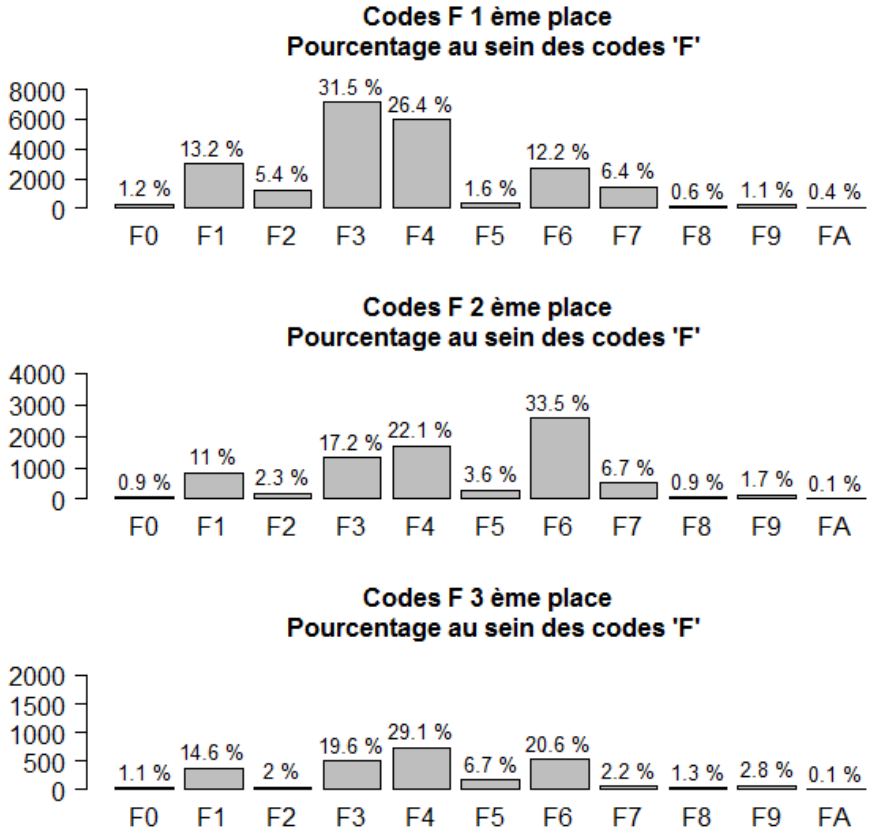


FIGURE 5.15 – Histogramme des codes "F" en fonction de leur place dans le diagnostic

Les codes F étaient les seuls pour lesquels regarder les deux premières lettres avait du sens, car ils sont les seuls possédant des sous-catégories bien définies par le premier chiffre. Focalisons-nous maintenant uniquement sur le code principal, donc le premier, et observons-le plus précisément, c'est-à-dire, tenons compte de ses 3 caractères.

Nous ne gardons que les codes ayant au moins 300 patients concernés, afin de ne pas alourdir les graphiques. Malgré tout, les pourcentages seront faits au sein de tous les codes, donc il est normal que nous n'ayons pas 100% au total de tous les bâtons. Dans un premier temps, nous allons séparer les codes F et Z. Nous obtenons pour le code F, la FIGURE 5.16 avec à la TABLE 5.5 l'intitulé de chaque code repris dans l'histogramme.

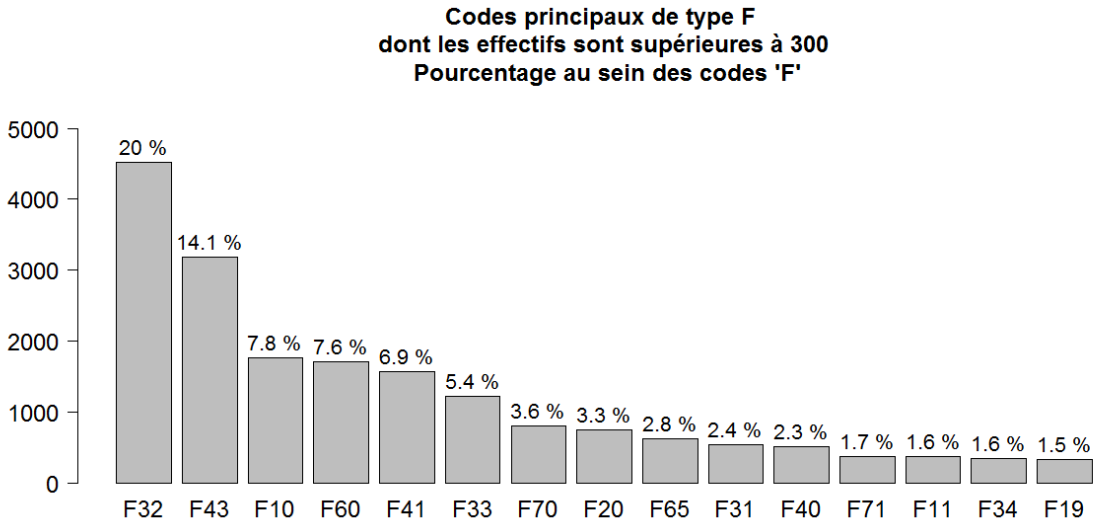


FIGURE 5.16 – Histogramme des codes F en diagnostic principal

Code	Intitulé
F32	Épisodes dépressifs
F43	Réactions à un facteur de stress important, et troubles de l'adaptation
F10	Troubles mentaux et troubles du comportement liés à l'utilisation d'alcool
F60	Troubles spécifiques de la personnalité et du comportement
F41	Autres troubles anxieux
F33	Trouble dépressif récurrent
F70	Retard mental léger
F20	Schizophrénie
F65	Troubles de la préférence sexuelle
F31	Trouble affectif bipolaire
F40	Troubles anxieux phobiques
F71	Retard mental moyen
F11	Troubles mentaux et troubles du comportement liés à l'utilisation d'opiacés
F34	Troubles de l'humeur
F19	Troubles mentaux et troubles du comportement liés à l'utilisation de substances psychoactives multiples et troubles liés à l'utilisation d'autres substances psychoactives

TABLE 5.5 – Les principaux codes F (troubles mentaux et troubles du comportement)

Nous observons sur cette FIGURE 5.16 qu'un cinquième des codes F correspondent à des épisodes dépressifs. Nous avons ensuite 14.1% de "Réactions à un facteur de stress important, et troubles de l'adaptation". Le suivant est un code de la catégorie "F1", ce qui est surprenant puisqu'il semblait y avoir plus de F3,F4 et F6. Malgré tout, quand on regarde le code complet, le F10, "Troubles mentaux et troubles du comportement liés à l'utilisation d'alcool" passe devant les F6 avec 7.8%. Celui-ci est tout de même suivi de près du code F60 avec 7.6%, ou "Troubles spécifiques de la personnalité et du comportement". Viennent alors les autres troubles anxieux et les épisodes dépressifs récurrents et, ensuite, tous les codes qui ne sont pas beaucoup représentés.

Pour les codes Z, nous avons le graphique correspondant à la FIGURE 5.17, suivi de la TABLE 5.6 contenant les intitulés ainsi que des exemples. Nous constatons qu'au sein des codes de type "Z", nous en avons un qui revient plus souvent que les autres. En effet, 46.4% de ces codes sont des "Z63", correspondant à "Autres difficultés liées à l'entourage immédiat, y compris la situation familiale". On ne sait pas exactement ce qui est repris dans cette catégorie, puisque ce sont les "Autres" difficultés, mais nous avons quelques exemples, tels qu'un décès dans la famille, des difficultés avec le conjoint,... Le deuxième code de type "Z" le plus présent ne comprend que 14.7% et est "Difficultés liées à l'environnement social". Nous avons ensuite les "Difficultés liées à d'autres situations psychosociales", les " Difficultés liées à une enfance malheureuse", puis la "Mise en observation et examen médical pour suspicion de maladies", et ainsi de suite.

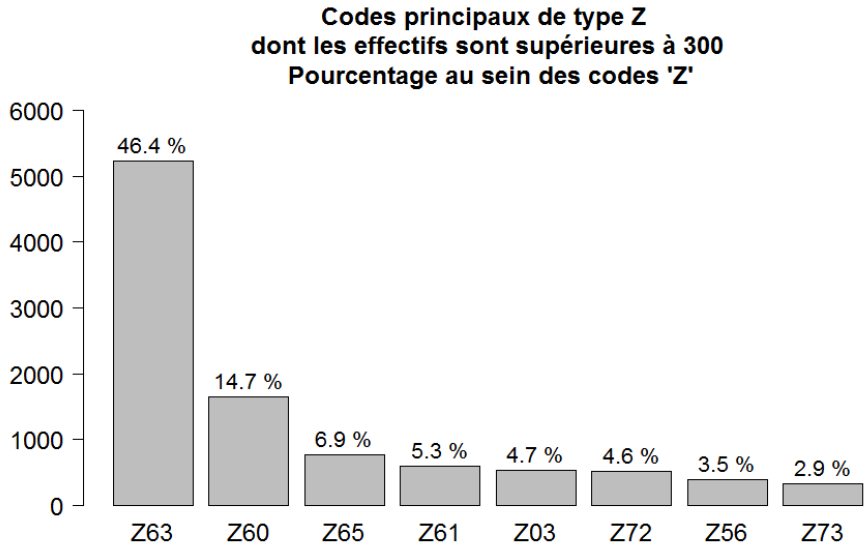


FIGURE 5.17 – Histogramme des codes Z en diagnostic principal

Code	Intitulé	Exemples
Z63	Autres difficultés liées à l'entourage immédiat, y compris la situation familiale	Décès d'un membre de la famille, difficultés dans les rapports avec le conjoint,..
Z60	Difficultés liées à l'environnement social	Solitude, difficultés d'adaptation à une nouvelle étape de vie, ...
Z65	Difficultés liées à d'autres situations psychosociales	Condamnation, emprisonnement, arrestation, victime d'un crime, ...
Z61	Difficultés liées à une enfance malheureuse	Privation de relation affective pendant l'enfance, départ du foyer, difficultés liées à de possibles sévices sexuels, ...
Z03	Mise en observation et examen médical pour suspicion de maladies	Mise en observation pour suspicion d'un trouble mental ou d'un trouble du comportement
Z72	Difficultés liées au mode de vie	Utilisation de drogues, d'alcool, régime et habitudes alimentaires inadéquates, ...
Z56	Difficultés liées à l'emploi et au chômage	
Z73	Difficultés liées à l'orientation de son mode de vie	Burnout, manque de repos ou de loisirs,...

TABLE 5.6 – Exemples pour les codes Z (troubles mentaux et troubles du comportement)

Une étape encore intéressante à développer consiste à reprendre l'ensemble des codes principaux des deux types simultanément. Nous savons que les codes "F" reviennent plus souvent que les "Z" pour les codes principaux. Le but du paragraphe suivant est de voir quels diagnostics principaux reviennent le plus souvent. Pour faire un histogramme de ce type, nous allons devoir n'afficher que les plus représentés, car il existe une multitude de codes différents. Ici, nous décidons de n'afficher que les codes concernant plus de 2% des effectifs. Les pourcentages sont malgré tout calculés sur l'ensemble des codes principaux. Cette illustration se trouve à la FIGURE 5.18.

Nous observons que, même si les codes "F" sont dans l'ensemble plus présents, le code le plus utilisé en première position est de type "Z", avec 15.1% des patients. En effet, c'est le "Z63", correspondant aux "Autres difficultés liées à l'entourage immédiat, y compris la situation familiale". Les quatre codes suivants sont de type "F", avec les "épisodes dépressifs" pour 13% des nouveaux patients. L'analyse des codes "F" et "Z" séparément a déjà été faite plus haut, donc l'intérêt de ce graphique est uniquement de voir comment ces codes s'emboitent.

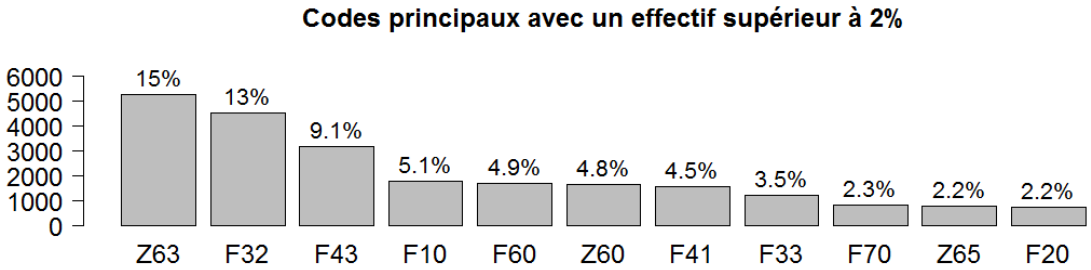


FIGURE 5.18 – Diagnostic principal (eff>700)

Nous pouvons également observer les codes de type différent de F et Z, mais ils sont déjà fort minoritaires. Nous ne garderons que ceux ayant des effectifs supérieurs à 50. Nous obtenons la FIGURE 5.19, avec la TABLE 5.7. Nous constatons que 16.2% des codes principaux de ce type sont des maladies non spécifiées. Ensuite, 13.9% sont des "Problèmes liés à l'abus ou la négligence". Les trois derniers qui reviennent pour plus de 50 codes généraux représentent plus ou moins la même proportion et sont l'épilepsie, l'"Intoxication volontaire par l'alcool et exposition à l'alcool" et l'"Intoxication volontaire par des stupéfiants et des psychodysleptiques (hallucinogènes), non classés ailleurs, et exposition volontaire à ces produits". Remarquons que ces deux derniers font partie d'une même catégorie qui est l'intoxication volontaire.

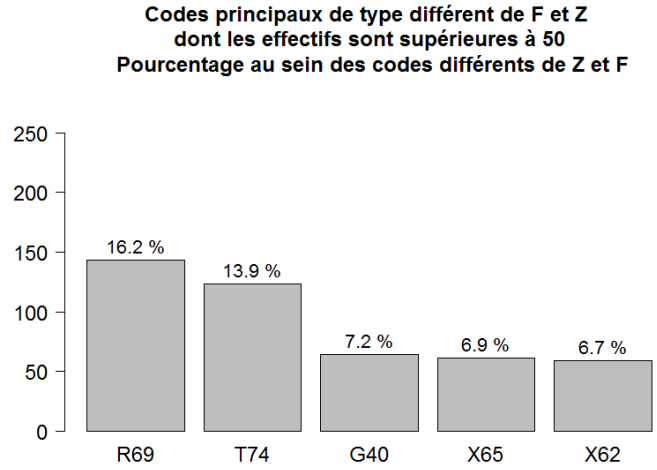


FIGURE 5.19 – Histogramme des codes différents de Z et F en diagnostic principal

Code	Intitulé
R69	Maladies non spécifiées
T74	Problèmes liés à l'abus ou la négligence
G40	Epilepsie
X65	Intoxication volontaire par l'alcool et exposition à l'alcool
X62	Intoxication volontaire par des stupéfiants et des psychodysleptiques (hallucinogènes), non classés ailleurs, et exposition volontaire à ces produits

TABLE 5.7 – Tableau des codes différents de "Z" et "F"

Passons maintenant à la prise en charge proposée au patient. Remarquons que l'on pouvait avoir jusqu'à 3 prises en charge proposées. Nous pouvons commencer par regarder globalement les prises en charge proposées, c'est-à-dire considérer toutes les propositions qu'elles soient en première, deuxième ou troisième position. Remarquons que pour la troisième proposition de prise en charge, nous avions 7762 erreurs d'encodage, c'est-à dire tous les patients de la province du Hainaut. Je ne les ai donc pas considérés dans les graphiques ci-dessous. Nous pouvons voir l'ensemble des prises en charge proposées à la FIGURE 5.20. Nous constatons que, pour presque la moitié des nouveaux patients, une thérapie est envisagée. De plus, 31.7% des personnes se voient proposer un accompagnement ou un soutien. Des informations sont données à 12.3% des patients et 11.4% n'ont aucune prise en charge immédiate proposée. Enfin, un traitement médical est suggéré à 10.2% des nouveaux patients.

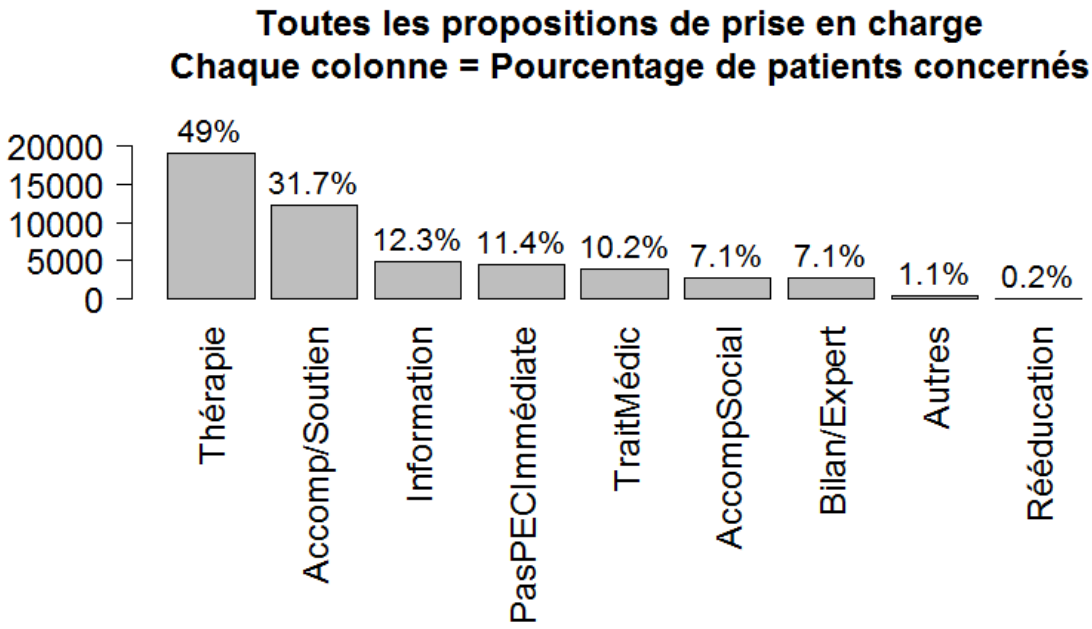


FIGURE 5.20 – Histogramme de l'ensemble des prises en charge proposées

Nous pouvons désormais séparer cette analyse par position. Nous avons ainsi les FIGURES 5.21. Pour rendre les graphiques plus clairs, je n'ai affiché que les pourcentages supérieurs à 5%.

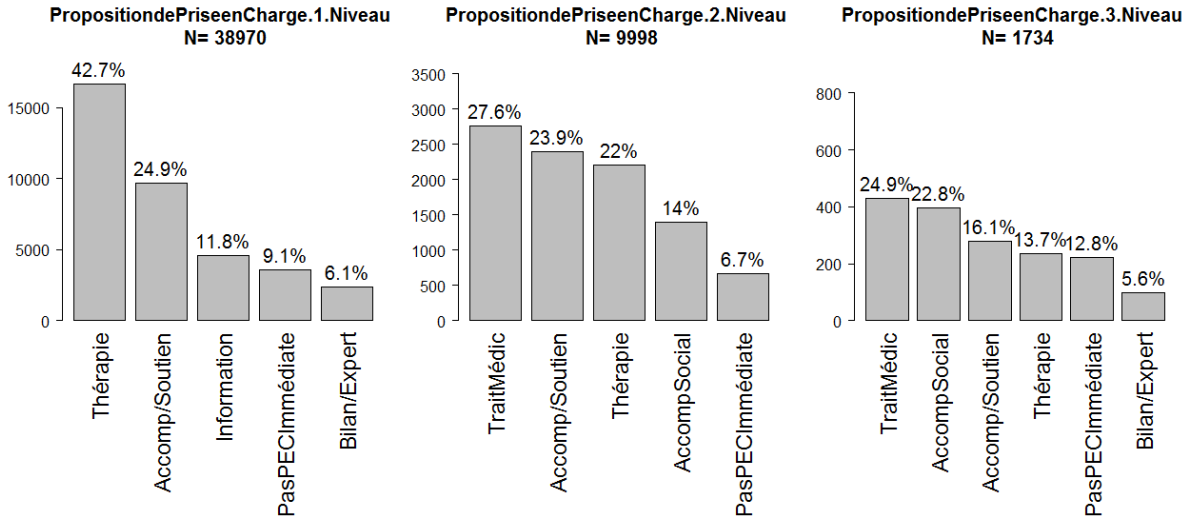


FIGURE 5.21 – Histogramme des prises en charge proposées par ordre d'importance

Nous pouvons voir que nous avons en position principale la "thérapie" qui est la plus présente, alors qu'en positions deux et trois, c'est le "traitement médical" qui est le plus courant. Pour la première place, le traitement médical ne figure pas sur l'histogramme donc représente moins de 5%. Pour la thérapie, qui est la plus présente en proposition principale, nous ne la retrouvons qu'en troisième position dans les deuxième et troisième prises en charge. Dans les trois cas, c'est un accompagnement ou un soutien, qui est le deuxième plus proposé. De plus, 9.1% des patients n'ont pas eu de proposition de prise en charge immédiate.

Chapitre 6

Analyse bivariée : dépendances et description

Étant donné la quantité d'informations et d'individus dans notre base de données, il ne sera pas faisable de représenter et d'analyser toutes les associations possibles de variables. Par conséquent, nous avons dans un premier temps identifié les associations qui seraient potentiellement intéressantes avec le groupe de travail. Cela reprenait un mélange de variables "profils" et de variables concernant la thérapie. Nous pourrons associer des variables de types différents ou celles concernant la thérapie ensemble. Dans le cadre de cette étude, nous ne nous intéresserons pas aux dépendances possibles entre plusieurs variables "profils".

Afin de restreindre les études développées plus en détails, nous allons d'abord déterminer les variables qui sont dépendantes. Par la suite, nous identifierons les types de dépendances et les illustrerons par des histogrammes.

6.1 Explications théoriques du Chi-Carré

Pour cela, nous effectuons un test d'hypothèses à l'aide d'un Chi-Carré. Le but des tests d'hypothèses est de déterminer, parmi deux hypothèses, laquelle est la plus probable. Dans notre cas, les deux hypothèses de base sont les suivantes :

- H0 : les variables sont indépendantes ;
- H1 : les variables ne sont pas indépendantes (corrélation possible).

La méthode du Chi-Carré consiste à quantifier l'écart entre le tableau des fréquences observées et celles qu'on aurait observé dans le cas d'une parfaite indépendance avec les mêmes fréquences marginales. Afin d'illustrer le fonctionnement de ce processus, prenons un exemple. Nous allons considérer la table des fréquences entre le type de dossier et le sexe. Nous voyons à la TABLE 6.1 le tableau de contingence de ces deux variables pour nos données.

Sexe-TypeDossier	Couple	Famille	Individu	Total
F	779	549	21 979	23 307
M	825	375	17 125	18 325
Total	1604	924	39 104	41 632

TABLE 6.1 – Exemple - contingence observée entre le type de dossier et le sexe

Nous avons en bleu et en rouge les fréquences marginales de chaque variable. L'étape suivante, pour savoir si les deux variables sont dépendantes ou non, consiste à calculer le tableau théorique qui aurait été observé en cas de parfaite indépendance. Pour cela, nous gardons les fréquences marginales, afin de s'assurer que les variables vues séparément auraient gardés les proportions observées. Ensuite, s'il y avait une réelle indépendance, il faudrait que les combinaisons soient distribuées de sorte à garder les proportions de chaque variable. Remarquons que nous n'avons pas considéré les "DM" et que, par conséquent, nous n'avons ici au total que 41 632 individus.

Calculons la première cellule du tableau théorique. Les 1604 dossiers de couples devraient être distribués de sorte que le sexe respecte les proportions totales. Par conséquent, il faudrait $\frac{23307}{41632} = 55.98\%$ de femmes et le reste d'hommes. Nous aurons donc dans la première case $1604 \cdot \frac{23307}{41632} = 897.97$ femmes avec un dossier de couple. En continuant à créer la table de contingence théorique de façon similaire, nous obtenons la TABLE 6.2 qui conserve bien les fréquences marginales.

Sexe-TypeDossier	Couple	Famille	Individu	Total
F	897.97	517.29	21 891.74	23 307
M	706.03	406.71	17 212.26	18 325
Total	1604	924	39 104	41 632

TABLE 6.2 – Exemple - contingence théorique entre le type de dossier et le sexe

Maintenant que nous avons notre table, notée Obs, et celle qu'on aurait en cas d'indépendance parfaite, notée Th, il reste à décider si ces contingences sont suffisamment proches l'une de l'autre. Pour faire cela, nous allons calculer pour chaque cellule du tableau l'écart de la manière suivante :

$$\chi^2_{ij} = \frac{(Obs_{ij} - Th_{ij})^2}{Th_{ij}}$$

Nous pouvons calculer cette quantité, appelée Chi-Carré et notée χ^2, pour chaque cellule et obtenir la TABLE 6.3. Le fait de calculer d'abord ce χ^2 pour chaque cellule nous permet de voir laquelle a la plus grande influence et est donc la plus éloignée de l'indépendance.

Sexe-TypeDossier	Couple	Famille	Individu
F	15.76	194	0.35
M	20.05	2.47	0.44

TABLE 6.3 – Exemple - χ^2_{ij} entre le type de dossier et le sexe

Il ne nous reste plus qu'à sommer toutes les cases du tableau pour obtenir le véritable χ^2 du test, ici 41.01. Le logiciel R calcule cette valeur automatiquement à l'aide de la fonction "chisq.test". Il nous renvoie alors le χ^2, ainsi qu'une quantité appelée "p-valeur". La p-valeur, notée p_v est la probabilité, sous l'hypothèse d'indépendance H0, d'observer l'échantillon récolté ou un pire, c'est-à-dire encore plus éloigné de celui correspondant à une indépendance. Cette p-valeur peut être également définie à l'aide d'une table de référence reprenant les différentes évaluations du Chi-Carré et de ces p-valeurs associées.

Pour trouver la p-valeur avec une table de référence, il faut connaitre le χ^2, ainsi que le degré de liberté du test. Celui-ci, noté n, correspond au nombre de colonnes moins 1, multiplié par le nombre de lignes moins 1. Dans notre exemple, nous avons 2 lignes et 3 colonnes, donc le degré de liberté vaut $(2-1)(3-1) = 2$. La FIGURE 6.1 contient une partie de la table de référence du Chi-Carré. La ligne qui nous intéresse est celle de degré de liberté égal à 2. Ensuite, le q correspond à notre p-valeur, soit à la probabilité que la valeur du χ^2 soit plus grand ou égal à celle calculée. Remarquons que, plus le χ^2 est grand, plus la p-valeur est petite. Nous constatons que notre 41.01 est tellement grand qu'il serait en dehors de la table. Par conséquent, sa p-valeur associée sera inférieure à 0.005. R confirme notre calcul en ressortant une p-valeur de l'ordre de 10^{-9}.

q	0,995	0,99	0,975	0,95	0,9	0,8	0,7	0,6	0,5	0,4	0,3	0,2	0,1	0,05	0,025	0,01	0,005	q
n																		n
1	0,000	0,000	0,001	0,004	0,016	0,064	0,148	0,275	0,455	0,708	1,074	1,642	2,706	3,841	5,024	6,635	7,879	1
2	0,010	0,020	0,051	0,103	0,211	0,446	0,713	1,022	1,386	1,833	2,408	3,219	4,605	5,991	7,378	9,210	10,597	2
3	0,072	0,115	0,216	0,352	0,584	1,005	1,424	1,869	2,366	2,946	3,665	4,642	6,251	7,815	9,348	11,345	12,838	3

FIGURE 6.1 – Table de référence du Chi-Carré [1]

Si la p-valeur est grande, c'est qu'il y a une grande probabilité d'obtenir ces données, si les variables sont indépendantes. Dans le cas contraire, c'est qu'il n'y a pas beaucoup de chances de tomber sur cet échantillon si les variables sont indépendantes. Dans ce cas, on rejette l'hypothèse H0 au profit de H1. Pour pouvoir formuler une décision finale, il est nécessaire d'avoir un seuil d'acceptation. Nous choisirons $\alpha = 0.05$, donc :

- $p_v < \alpha$: nous rejetons H0 (au profit de H1) $\rightarrow$ dépendance
- $p_v \geq \alpha$: nous ne rejetons pas H0 $\rightarrow$ indépendance

Remarquons que, pour que le test soit valable, il faut que les cases de la table des fréquences contiennent chacune au moins 5 personnes. Il est nécessaire pour les colonnes possédant trop de réponses différentes de faire des classes afin de ne pas avoir de modalités comprenant trop peu de données. Nous allons donc faire des classes pour certaines variables. Ces groupements seront expliqués lorsqu'ils seront utilisés.

6.2 Application de tests Chi-Carré à nos données

Dans le cadre de ce mémoire, procéder à un test d'hypothèses sur l'ensemble des données renvoie systématiquement une p-valeur très proche de 0 et donc, tout serait dépendant. Cela est dû à la grande quantité de données. En effet, nous avons des données très précises avec les 41 916 lignes, et cela cause des écarts entre les données théoriques et observées beaucoup trop grands. Nous ne pouvons pas considérer que tout est dépendant, mais il faut trouver un moyen de contourner le problème de surplus d'informations.

Dans un premier temps, nous avons pensé à considérer chaque année d'enregistrement séparément. De cette manière, nous aurions moins de données pour chaque test d'indépendance entre deux variables, et nous pourrions par la suite faire la moyenne des p-valeurs obtenues. Nous avons fixé la première variable à "Sexe" et regardé sa dépendance avec plusieurs autres données. Cela nous donne la Table 6.4. Pour créer cette table, j'ai décidé de garder 3 décimales. De plus, la nature de la démarche est considérée sans la modalité "Autre" qui était sous-représentée et pour l'origine de la démarche, nous avons dû joindre "invalidité" à "handicap", ainsi que "PetiteEnfance" et "3emeAge" à "Autres".

variable	2008	2009	2010	2011	moyenne
TypeDossier	0.034	0.003	0.016	0	0.013
NatureDemarche	0	0	0	0	0
OrigineDemarche	0	0	0	0	0
DemandeConsultant	0	0	0	0	0
Motifs	0	0	0	0	0

Table 6.4 – Tests d'indépendance avec le sexe par année d'enregistrement

Pour tous les essais qui ont été réalisés avec cette technique, la grande majorité nous donnait à nouveau des dépendances. Nous avons décidé que cette méthode n'était pas très concluante. De plus, elle nous empêchait de tester formellement la dépendance entre les variables et l'année d'enregistrement.

Notons que les tests d'hypothèses ne sont pas tout à fait adaptés à notre situation. En effet, un test d'hypothèses est une étude inférentielle des données. Cela signifie qu'il s'applique généralement à une partie de la population, appelée échantillon, et qu'il essaye de tirer des conclusions sur les dépendances dans la population complète. Or, dans

notre cas, nous avons les informations sur l'ensemble des patients désirés, donc nous n'avons pas un échantillon, mais un recensement.

Généralement, on ne procède que très rarement à des recensements, car cela nécessite d'atteindre l'ensemble de la population concernée et coûte donc cher en investissement pour l'enquête. On se contente donc d'échantillons, en prenant garde à la façon dont ils sont choisis. De plus, atteindre l'ensemble des personnes prend du temps, et l'échantillon permet d'en gagner.

Nous allons donc procéder à un échantillonnage aléatoire. La suite de l'explication se base sur le livre "Sampling of Populations" [18]. Il existe deux types d'échantillons : ceux de probabilité et ceux sans probabilité. Un échantillon de probabilité consiste à donner à chaque élément de la population totale la même probabilité d'être dans l'échantillon. Ce genre d'échantillonnage nous permet d'estimer des caractéristiques sur la population totale sans biais. Pour prendre un échantillon aléatoire simple, il faut prendre aléatoirement n individus au sein des N constituant l'ensemble de la population. Pour ce faire, on donne à chaque individu un numéro unique entre 1 et N et on génère n nombres aléatoires différents dans cet intervalle. De cette manière, on obtient un échantillon de probabilité, puisque chaque individu a eu la même probabilité d'être choisi (soit $\frac{n}{N}$).

Remarquons que des échantillons de probabilité sont très difficiles à récolter en pratique. En effet, il faut que chaque individu ait exactement la même probabilité d'être dans l'échantillon. Par conséquent, il faut un échantillon qui couvre une grande zone géographique et ne soit pas du tout biaisé. Par exemple, demander aux personnes de remplir le questionnaire à la sortie d'une école implique que les étudiants ont plus de chances de se trouver dans l'échantillon. Un autre exemple, si nous envoyons l'enquête par mail, les personnes qui ne possèdent pas d'ordinateur ne seront pas représentées et nous risquons d'avoir un échantillon ne comprenant que très peu de personnes âgées. Demander un échantillon de probabilité demande un énorme travail lors de l'enquête et beaucoup de réflexion afin d'être certains que l'échantillon est valable.

Dans notre cas, cela n'est pas compliqué puisque nous avons les informations sur la population entière des nouveaux arrivants en service de santé mentale. Nous avons justement trop d'informations. Nous sommes sûrs d'avoir un échantillon de probabilité et pourrons utiliser les tests associés.

Comme vu au cours d'introduction aux logiciels statistiques [25], un échantillon est valable avec une certaine marge d'erreur notée l si le nombre d'individus de l'échantillon, noté n, dépasse une valeur déterminée par la marge d'erreur et la taille de la population totale N. Cette limite est définie par la formule suivante :

$$n \geq \frac{(1.96)^2 \cdot N}{(1.96)^2 + l^2 \cdot (N-1)}$$

Dans notre cas, nous avons 41 916 individus, donc $N = 41916$. Afin d'avoir une idée de comment évolue la taille de l'échantillon lorsque le seuil d'incertitude augmente, nous le calculons pour différents seuils. On obtient le tableau de la TABLE 6.5.

l	2%	3%	4%	5%	6%	8%
n	7854	3875	2271	1483	1041	592

TABLE 6.5 – Taille minimum de l'échantillon en fonction du dergé d'incertitude

Nous choisirons un seuil d'incertitude de 3%. Un seuil plus petit aurait pu convenir mais, si nous prenons un trop petit échantillon, nous n'aurons pas assez d'individus. Nous tomberions dans le problème inverse, où chaque catégorie est sous-représentée et il nous serait impossible de tirer des conclusions. C'est pour cette raison que 3875 nous semble un bon choix. De cette manière, les problèmes liés au surplus d'informations seront évités avec un petit seuil d'incertitude et sans avoir trop peu de données.

Nous avions pensé toujours utiliser l'échantillonnage séparément par année d'enregistrement, mais cela nous empêche de calculer si des dépendances entre l'année et d'autres données sont présentes.

Dans notre cas, puisque nous avons les données pour l'ensemble de la population, nous pouvons nous permettre de faire l'analyse pour plusieurs échantillons et de calculer la moyenne des p-valeurs. Cette méthode nous permet de restreindre les problèmes qui pourraient survenir en cas d'échantillon peu représentatif de la population. Par conséquent, nous prendrons plusieurs échantillons de 3875 individus au sein de toutes les données. Nous prendrons 100 échantillons différents avant de faire la moyenne.

Nous allons, dans ce chapitre, tenter de trouver ce qui est corrélé avec le type de problème rencontré par le nouveau patient, ainsi que la solution proposée. Pour ce faire, nous considérerons, comme variable explicative, les données socio-démographiques en deux parties : tout d'abord, les données concernant le point de vue du patient de la consultation ; et ensuite les données émises par le praticien, comprenant les diagnostics et les prises en charge proposées. Pour faire ces tests, nous avons dû grouper des catégories, car certaines d'entre elles ne possédaient pas suffisamment d'effectifs. Dans certains cas, j'ai également retiré certaines catégories très peu représentées.

6.3 Consultation

Pour les variables concernant la consultation, nous n'avons pas considéré "Autre" dans la nature de la démarche ; nous avons groupé "invalidité" et "handicap", puis "PetiteEnfance", "3EmeAge" et "Autre" dans l'origine de la démarche et les motifs "Autre" et "SansMotifs" sont également considérés ensemble. Passons aux variables socio-démographiques. Nous n'avons considéré que les provinces wallonnes. Le mode de vie ne conserve que "familial" et "seul". L'âge est considéré avec ses classes en retirant les classes "HorsLimites" et "PersAgees". Les catégories professionnelles peu présentes sont groupées dans "Autre" de sorte à n'avoir plus que "Employé", "Ouvrier", "SansProf" et "Autre". Pour terminer, les ressources rassemblent "EntourSocial", "EntourProf" et "Amis" et retirent "Autre".

Les p-valeurs obtenues pour les variables concernant le point de vue du patient sur la consultation, se trouvent dans la Table 6.6. Rappelons que nous faisons la moyenne des p-valeurs sur 100 échantillons. Il peut arriver que, pour certains échantillons, nous ayons trop peu d'effectifs dans une catégorie. Nous avons ajouté une étoile aux p-valeurs lorsque cela s'est produit et que des "warnings" sont apparus. Ensuite, les p-valeurs signifiant qu'il y a une dépendance se trouvent en couleur : bleu si il y a également une étoile et rouge sinon.

	AnneeEnreg	Sexe	ModeVie	Age	Province	CatProf	Ress
NatDémarche	0.36	0.00	0.00	0.25	0.01	0.00	0.01
OrigDémarche	0.34	0.00	0.00	0.00	0.00*	0.00*	0.00*
TypeDossier	0.34	0.21	0.00	0.07	0.00	0.01	0.49
DemConsult	0.36	0.00	0.24	0.00*	0.00*	0.00*	0.00*
MotifsNiv	0.47	0.03	0.00	0.01	0.12	0.00	0.00

Table 6.6 – Dépendances avec les variables de la consultation concernant le patient

Cette Table 6.6 nous montre qu'aucune variable ne dépend de l'année de l'enregistrement, donc que le comportement ne change pas fort entre les différentes années. Ensuite, on peut voir que les hommes et les femmes ont des tendances différentes pour chaque donnée, sauf pour le type de dossier qui est indifférent au sexe du patient. Le mode de vie, lui, a une dépendance avec toutes les variables hormis la demande du consultant. Donc, le mode de vie du patient a une corrélation avec le motif qui amène le patient, l'origine et la nature de la démarche, ainsi que le type de dossier. De la même manière, nous pouvons voir les variables qui sont sensibles à la tranche d'âge, la province, la catégorie professionnelle et les ressources.

Au sein de ses corrélations, certaines intéressaient plus particulièrement l'OWS. En effet, il a été demandé de faire une analyse des dépendances du sexe, de l'âge et du mode de vie. Nous allons représenter les corrélations avec ces trois variables socio-démographiques.

Sexe

Tout d'abord, commençons par analyser les différentes influences avec le sexe. La corrélation avec la nature de la démarche est représentée à l'aide d'un "mosaic plot". Le principe de ce genre de présentation est le suivant. Tout d'abord, la nature de la démarche est insérée de manière à ce que la largeur de chaque rectangle soit proportionnelle au nombre de personnes au sein de cette catégorie. On a donc, dans un premier temps, le graphique de la FIGURE 6.2, où on peut bien vérifier qu'on a plus de personnes orientées, et moins contraintes.

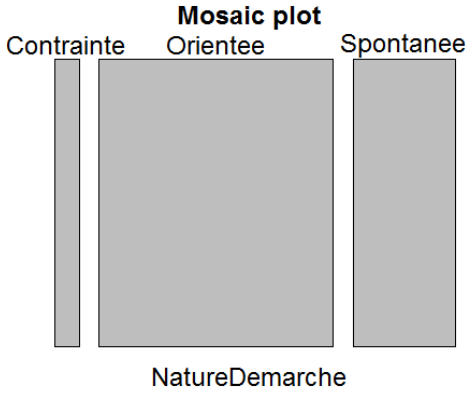


FIGURE 6.2 – Explication de la construction d'un mosaic plot

Ensuite, au sein de chaque catégorie, la répartition d'une autre variable, telle que le sexe dans notre cas, est représentée. Au final, nous obtenons le mosaic plot de la FIGURE 6.3. Cette illustration montre clairement que les personnes contraintes sont majoritairement des hommes. En effet, les hommes sont présents à presque 90% au sein des natures contraintes. Ensuite, les femmes représentent environ 60% des démarches spontanées et orientées. Nous avions presque 56% de femmes dans l'ensemble des données, donc les femmes sont plus susceptibles d'avoir ce genre de démarches.

Remarquons que nous avons effectué beaucoup de recherches sur les différentes façons de visualiser les dépendances. Nous avons choisi d'utiliser la représentation la plus visuelle, simple et intuitive. Lorsque ce graphique a été présenté à l'OWS, le groupe de travail a beaucoup apprécié cette découverte permettant une illustration claire de la table de contingence de deux données.

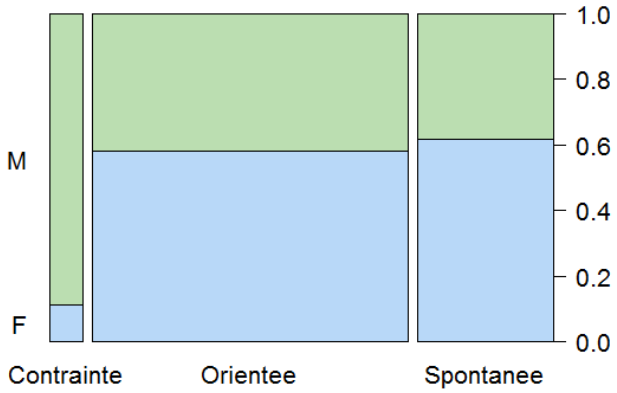


FIGURE 6.3 – Mosaic plot du sexe en fonction de la nature de la démarche

Nous allons faire ce genre de représentation pour les différentes corrélations à analyser. La suivante est l'origine de la démarche par sexe. Nous avions précédemment établi qu'il y avait un lien entre la nature et l'origine de la démarche. Nous pouvons donc bien nous imaginer que les hommes qui sont plus contraints le seront par la justice ou la police. La FIGURE 6.4 contient le mosaic plot correspondant. Nous confirmons que la justice et la police sont les origines principales des démarches des hommes. Les hommes sont également majoritaires dans l'origine de la démarche "handicap". Pour les autres modalités, les femmes sont les plus présentes, plus particulièrement pour le troisième âge.

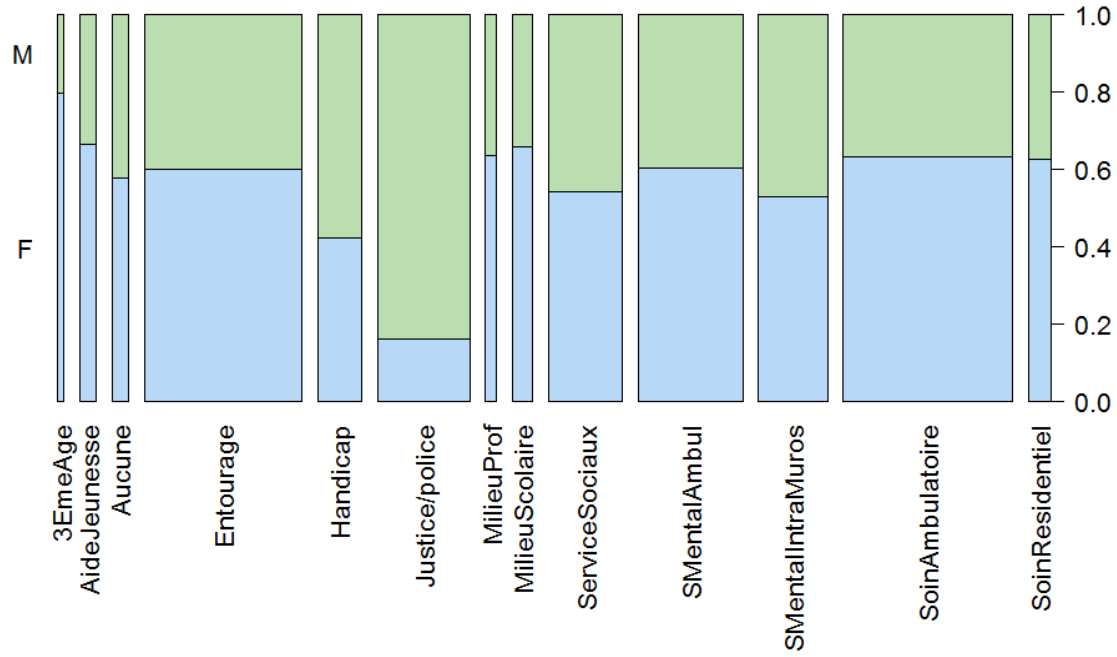


FIGURE 6.4 – Mosaic plot du sexe en fonction de l'origine de la démarche

Au niveau des demandes, la FIGURE 6.5 indique que les femmes attendent généralement un suivi, un conseil ou un avis, alors que les hommes demandent une attestation, un bilan ou une expertise. Pour les autres catégories, la différence est moins marquée, mais on peut quand même voir qu'en terme d'autres demandes et d'inscriptions ou de ré-évaluations AWIPH, les hommes sont plus demandeurs.

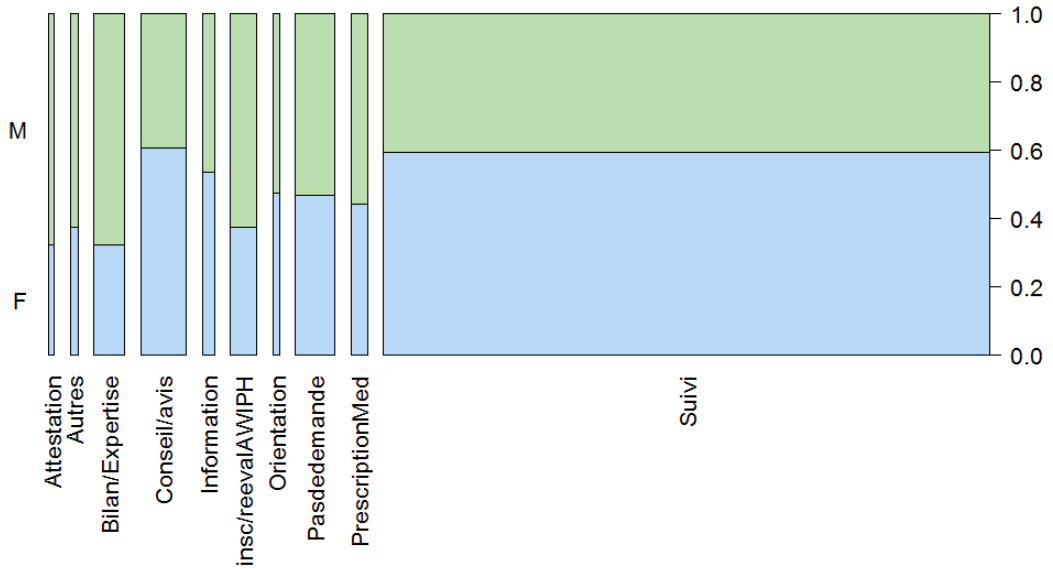


FIGURE 6.5 – Mosaic plot du sexe en fonction de la demande

La dernière donnée qui est corrélée avec le sexe est le motif de la consultation illustré à la FIGURE 6.6. L'élément le plus marquant de ce graphique est que plus de 90% des problèmes liés à des actes délictueux touchent les hommes.

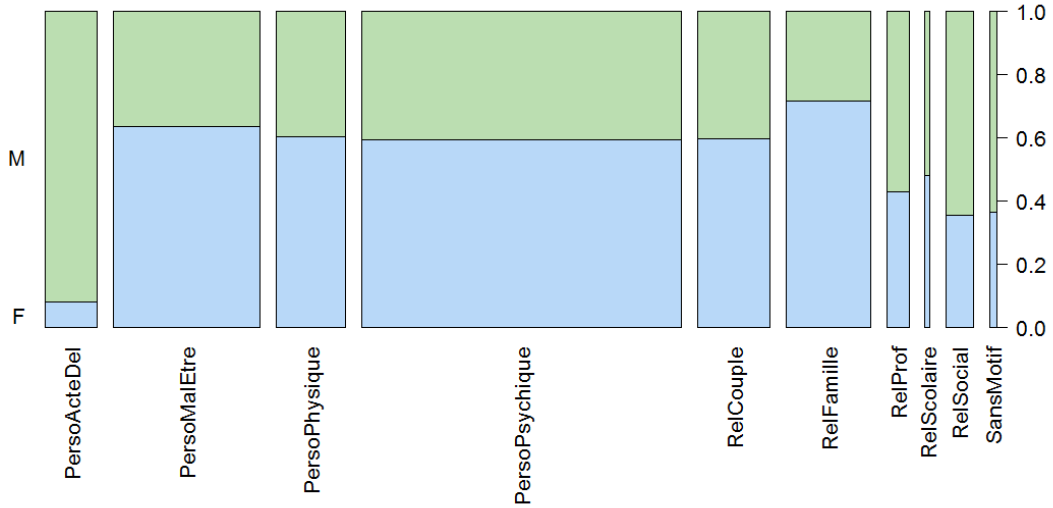


FIGURE 6.6 – Mosaic plot du sexe en fonction de la demande

Après en avoir discuté à l'OWS, nous avons réalisé que cela n'est pas très étonnant sachant que les prisons sont majoritairement occupées par des hommes. Les actes délictueux touchent clairement plus d'hommes que de femmes. Les problèmes de relations professionnelles ou sociales sont également plus présents chez les hommes. Par contre, les femmes sont plus concernées par les problèmes relationnels avec la famille.

Mode de vie

L'analyse des dépendances avec le sexe étant désormais terminée, nous pouvons nous focaliser sur le mode de vie. Remarquons que, afin que le mosaic plot soit lisible, j'ai rassemblé toutes les catégories sous-représentées dans "Autres". Par conséquent, la catégorie "Autre" contient "SDF", "Prison/DefSocial", "Institution" et "Communautaire". La première analyse, à la FIGURE 6.7, concerne la nature de la démarche. Nous observons que les personnes vivant seules ou dans un autre mode de vie sont proportionnellement plus représentées au sein de démarches contraintes. Cela pourrait s'expliquer par le fait que le mode de vie "Autre" contient des milieux dans lesquels les personnes n'ont pas nécessairement l'occasion de faire exactement de qu'elles veulent comme elles veulent, mais sont guidées. Pour les personnes seules, on peut s'imaginer qu'elles ont moins de personnes autour d'elles susceptibles de les orienter et de les soutenir dans une démarche spontanée. Au niveau des autres natures, le mode de vie "seul" est quasiment aussi présent dans les démarches spontanées et orientées, alors que le "familial" est plus présent dans les spontanées et les autres dans les orientées.

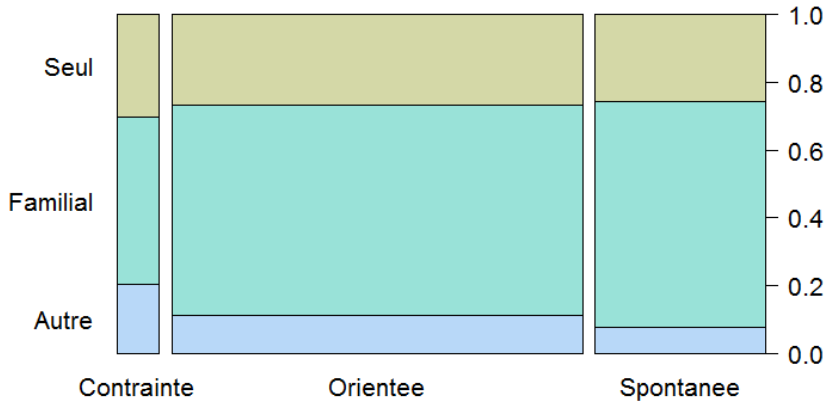


FIGURE 6.7 – Mosaic plot des modes de vie en fonction de la nature de la démarche

Le mosaic plot du mode de vie en fonction de l'origine de la démarche se trouve à la FIGURE 6.8. Les personnes vivant seules sont plus représentées au sein des personnes dont les services à l'origine de la démarche sont le troisième âge, les services sociaux, la santé mentale intra-muros et les soins résidentiels. Les autres modes de vie sont plus présents au sein du troisième âge, la justice ou la police, les services sociaux et le handicap. Cela est probablement dû au fait que les modes de vie présents dans "Autre" sont principalement ceux où la personne ne vit pas chez elle, mais dans une institution

ou un milieu communautaire. L'origine de leur démarche est donc au sein des réseaux présents autour d'eux. Les services ambulatoires, qu'ils soient de type santé mentale ou non, ainsi que l'entourage sont deux origines où l'on retrouve beaucoup de personnes vivant en famille. On a également les milieux scolaires et professionnels fort présents pour les modes de vie familiaux. Par contre, nous pouvons voir qu'au sein des troisièmes âges, il n'y a que très peu de personnes vivant dans un cadre familial.

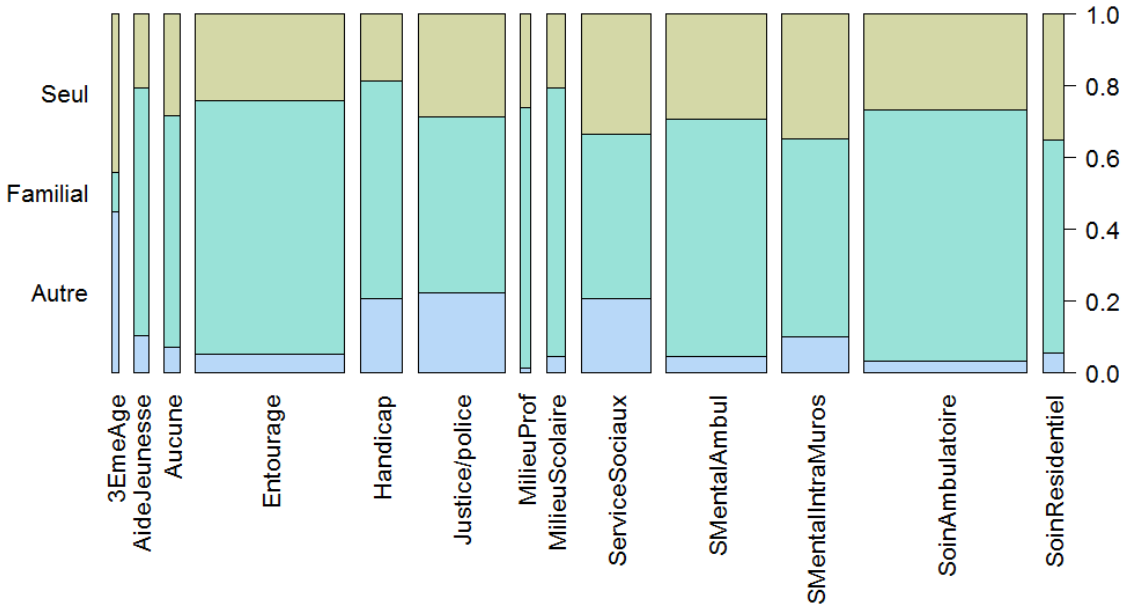


FIGURE 6.8 – Mosaic plot des modes de vie en fonction de l'origine de la démarche

Le type de dossier est également une variable montrant une dépendance au mode de vie. La FIGURE 6.9 décrit les combinaisons de ces deux données.

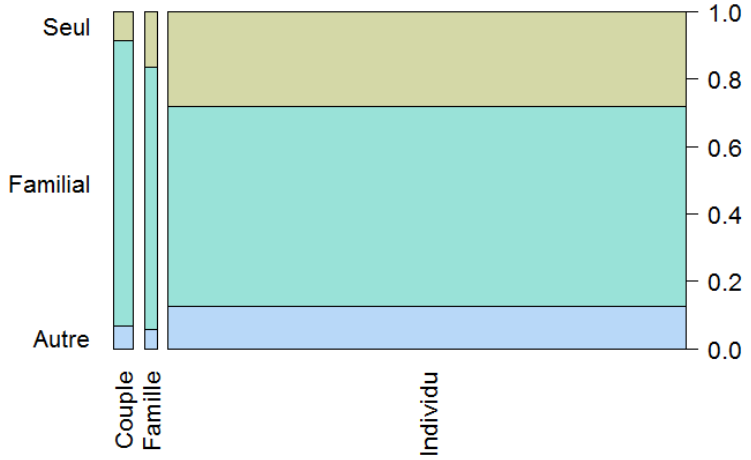


FIGURE 6.9 – Mosaic plot des modes de vie en fonction des types de dossier

Les personnes vivant seules ont une plus grande participation dans les dossiers de type "Individu", alors que les patients vivant dans un cadre familial sont plus représentés dans les dossiers de couple ou de famille. Cela semble assez intuitif, puisqu'en vivant seul on est moins susceptible d'avoir des problèmes de couple ou de famille, donc de se rendre à plusieurs à la consultation. Les autres modes de vie concernent principalement des dossiers "Individu".

Les analyses des correspondances entre le motif de la consultation et le mode de vie sont à la FIGURE 6.10. Cette illustration confirme que les patients vivant dans un milieu familial seront plus sujets à des problèmes relationnels de couple ou avec la famille. Les personnes habitant seules sont plus présentes dans tous les problèmes de type personnel, ainsi que dans les problèmes de relation sociale. Pour les autres modes de vie, ils sont majoritairement représentés dans les actes délictueux, ainsi que dans les relations sociales. Comme dit précédemment, cela s'explique par la composition des modes de vie inclus dans "Autre".

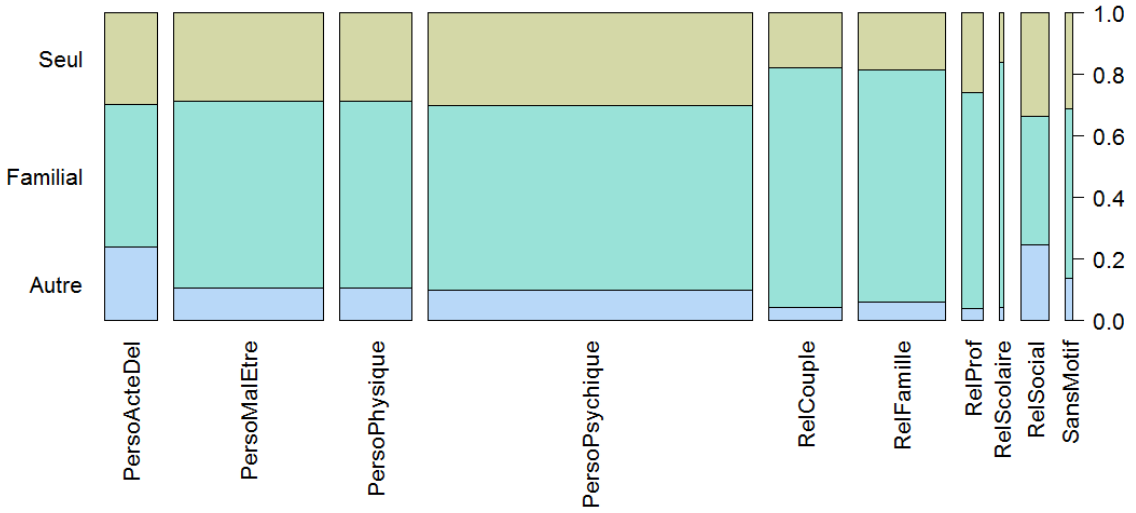


FIGURE 6.10 – Mosaic plot des modes de vie en fonction du motif de la première consultation

Âge

Nous pouvons nous concentrer désormais sur les dépendances avec l'âge. Afin de faciliter les représentations, nous n'avons pas considéré l'âge, mais la classe d'âge de chaque patient. Nous avions une première corrélation avec l'origine de la démarche. La FIGURE 6.11 indique que les jeunes sont plus touchés par les services "Handicap", ainsi que par les milieux scolaires. Il est très marqué et pas étonnant que les personnes âgées sont largement majoritaires au sein du troisième âge. Elles sont également fort présentes dans les soins résidentiels. Les adultes, entre 44 et 59 ans, ont une plus grande proportion de soins ambulatoires classiques ou de santé mentale et de soins de santé mentale intra-muros.

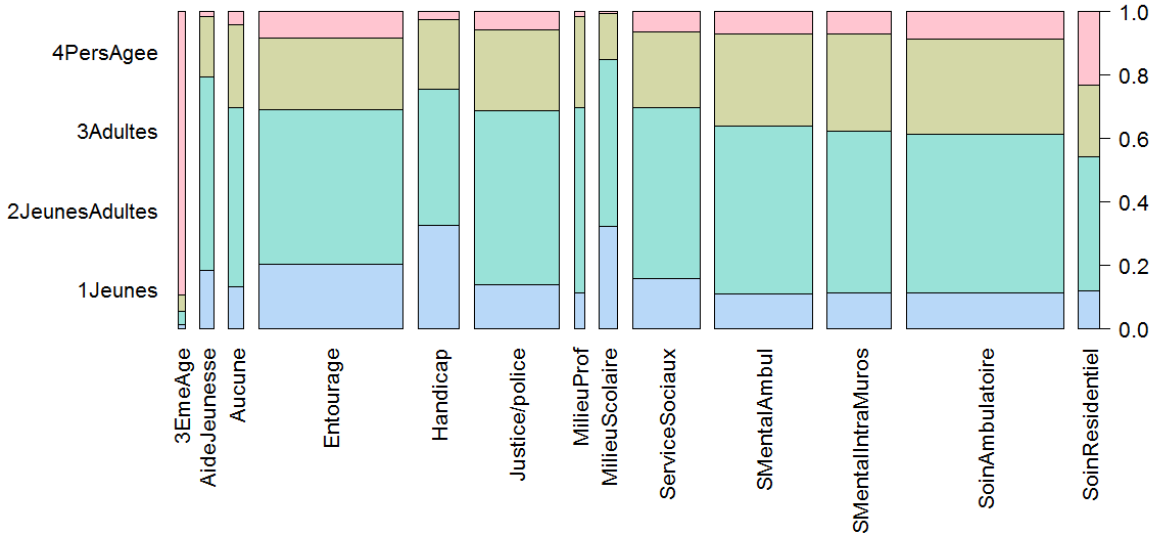


FIGURE 6.11 – Mosaic plot des âges en fonction de l'origine de la démarche

Remarquons que ce genre de graphique a des limites. En effet, pour les catégories centrales, il est moins aisé de tirer des conclusions lorsque nous avons beaucoup de modalités. Nous avons donc été regarder les pourcentages de chaque classe d'âge dans chaque modalité de l'origine avant de faire la FIGURE 6.12, où chaque modalité de l'origine représente 100% qui sont partagés entre les différentes classes d'âge.

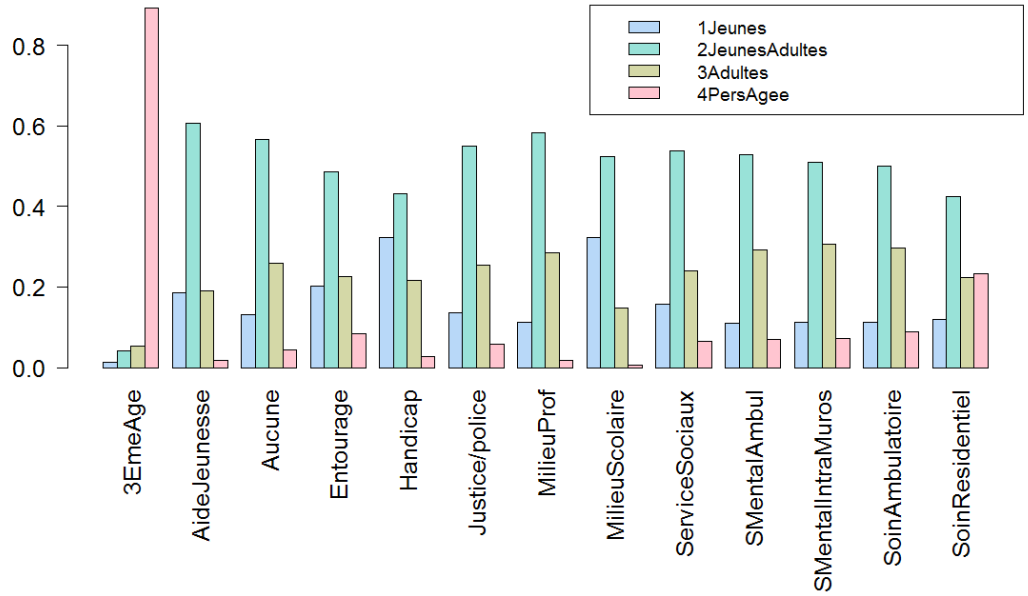


FIGURE 6.12 – Histogramme des âges en fonction de l'origine de la démarche

Cette nouvelle illustration confirme les observations précédentes. De plus, nous pouvons y voir que pour les soins résidentiels, les personnes âgées passent au-dessus des adultes, alors qu'elles n'étaient que peu représentées dans les autres origines, hormis le troisième âge.

Nous pouvons maintenant passer à l'analyse de l'âge avec la demande du consultant, à la FIGURE 6.13. Les personnes âgées demandent plus de conseils, d'avis ou d'informations et beaucoup d'entre elles n'ont aucune demande. Les jeunes sont plus demandeurs d'inscriptions et réévaluations "AWIPH", ainsi que de bilans et expertises, mais moins de prescriptions médicales. Les adultes ne semblent pas avoir de catégories vraiment privilégiées ou délaissées.

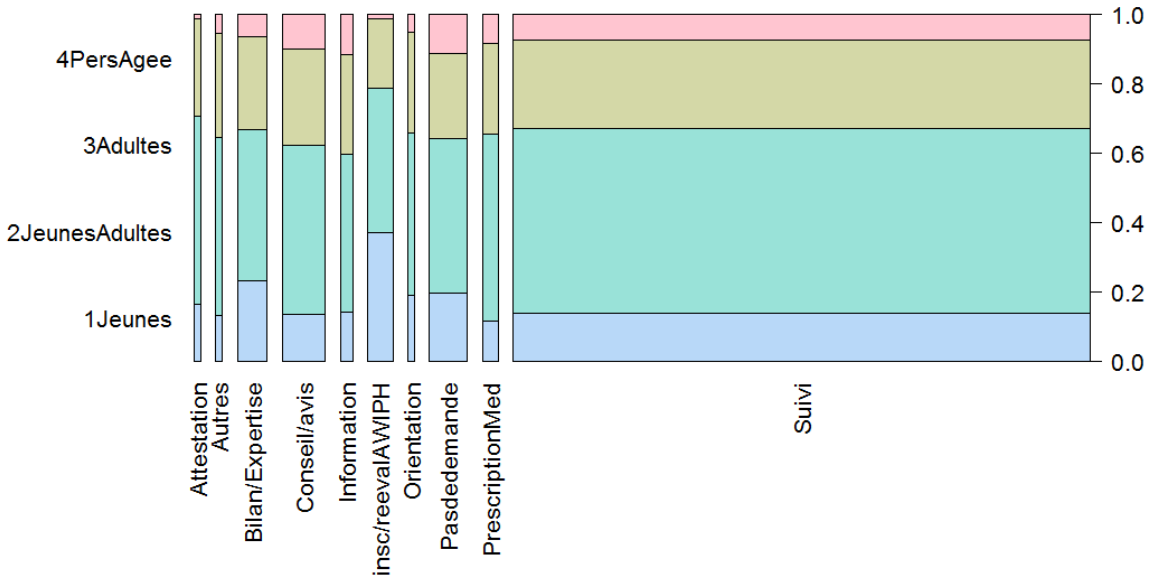


FIGURE 6.13 – Mosaic plot des âges en fonction de la demande

À nouveau, pour tirer des conclusions sur les adultes et les jeunes adultes, nous traçons l'histogramme de ces données, à la FIGURE 6.14. Les adultes sont légèrement plus présents dans les autres demandes et dans les orientations, et moins dans les inscriptions et réévaluations "AWIPH". Pour les jeunes adultes, de 25 à 44 ans, on peut voir qu'ils sont plus au sein des prescriptions médicales, des attestations et du suivi.

Cette illustration nous permet de voir que les jeunes et les jeunes adultes ont presque la même proportion des inscriptions et réévaluations "AWIPH", alors que dans toutes les autres demandes nous avons beaucoup plus de jeunes adultes que d'adultes. Par conséquent, les jeunes ont presque autant besoin de l'Agence Wallonne pour l'Intégration des Personnes Handicapées que les adultes. Par contre, cette demande est très faible au niveau des personnes handicapées.

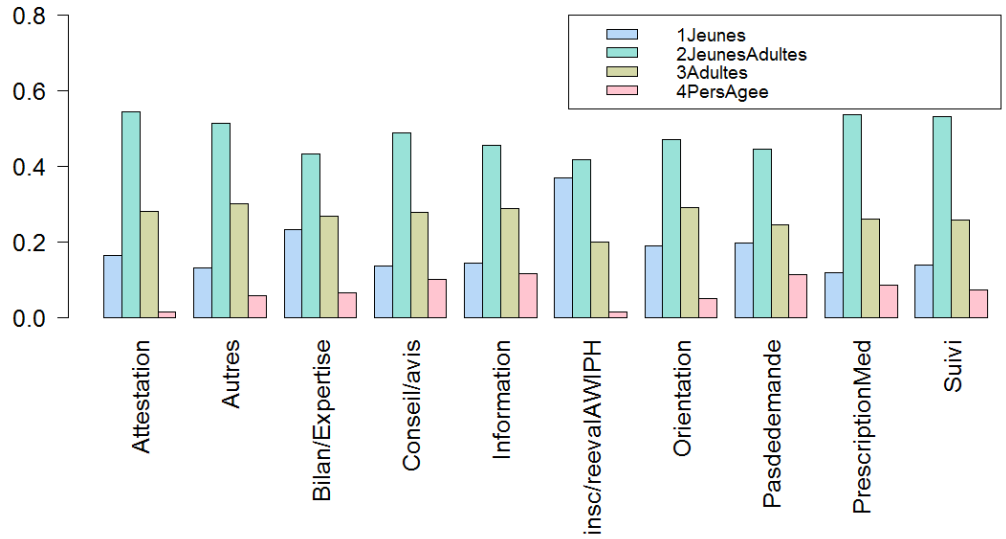


FIGURE 6.14 – Histogramme des âges en fonction de la demande

La dernière corrélation à développer, avant de passer à l'analyse du praticien, est l'âge avec la demande du consultant, à la FIGURE 6.15 et 6.16 . Nous avons décidé de tracer directement les deux sortes de représentations. Analysons dans un premier temps ce qui est visible sur la FIGURE 6.15. Les personnes âgées viennent plus à cause de problèmes physiques ou de relations sociales, tandis que la cause de la présence des jeunes est plutôt les problèmes de relations scolaires et professionnelles. Beaucoup de jeunes viennent également sans motif.

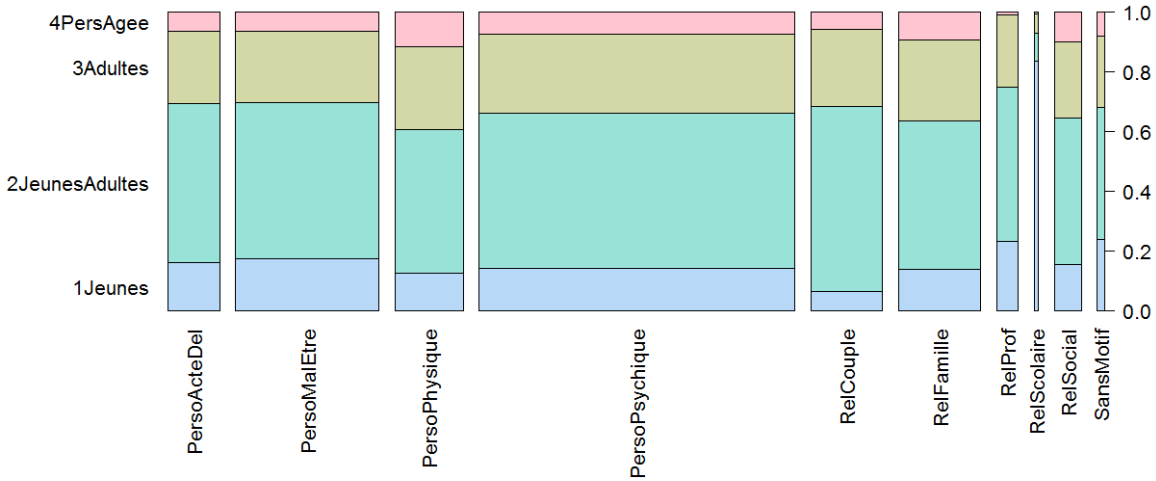


FIGURE 6.15 – Mosaic plot des âges en fonction du motif de la première consultation

Grâce à la FIGURE 6.16, nous pouvons voir que les problèmes de couple touchent bon nombre de jeunes adultes. De plus, les adultes n'ont pas de catégories où ils sont vraiment plus présents, ils sont juste très peu dans les problèmes relationnels scolaires. Ce motif a un histogramme particulier, puisque c'est le seul pour lequel la majorité est fort grande, plus de 80% de jeunes et de moins en moins d'effectif quand l'âge augmente. Cela est logique, car généralement, nous allons à l'école quand nous sommes plus jeunes.

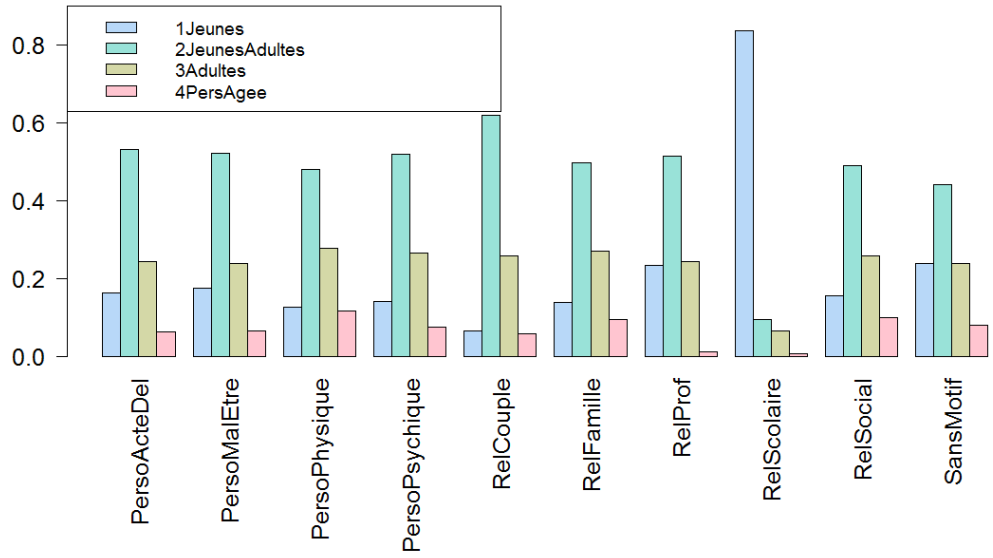


FIGURE 6.16 – Histogramme des âges en fonction du motif de la première consultation

Ceci termine l'analyse des corrélations de cette section.

6.4 Analyse du praticien

Nous pouvons désormais faire la table des p-valeurs des tests d'hypothèses concernant les mêmes données socio-démographiques, mais avec les variables concernant la participation du praticien à l'analyse, c'est-à-dire son diagnostic et ses propositions de prise en charge. Nous avons les résultats à la TABLE 6.7.

	AnneeEnreg	Sexe	ModeVie	Age	Province	CatProf	Ress
Code	0.33	0.00	0.00	0.00	0.00	0.00	0.00
PropPEC	0.29	0.00	0.01	0.00	0.00	0.00	0.00*

TABLE 6.7 – Dépendances avec les variables de la consultation déterminées par le praticien

Nous avons à nouveau en rouge les dépendances et en bleu les dépendances pour lesquels le test de chi-carré a exprimé un "warning" causé par des catégories sous-représentées. Nous pouvons voir que le diagnostic et les prises en charge proposées ne sont pas dépendantes de l'année d'enregistrement, mais bien de toutes les autres variables. Nous allons à nouveau tenter d'illustrer les dépendances avec le sexe, l'âge et le mode de vie. Rappelons que les codes "F" sont des troubles mentaux et du comportement, dont le premier chiffre indique dans quelle catégorie le code se trouve. Ces classes sont rappelées à la TABLE 6.8. Les codes "Z" sont les facteurs influant sur l'état de santé, donc ne sont pas un diagnostic en soi et les autres codes sont les diagnostics médicaux.

	Intitulé	Exemples
F0	Troubles mentaux organiques, y compris les troubles symptomatiques	Maladie d'Alzheimer
F1	Troubles mentaux et troubles du comportement liés à l'utilisation de substances psychoactives	Troubles mentaux et troubles du comportement liés à l'utilisation d'alcool ou de cocaïne
F2	Schizophrénie, trouble schizotypique et troubles délirants	Schizophrénie, trouble psychotique
F3	Troubles de l'humeur (affectifs)	Episode dépressif, trouble bipolaire
F4	Troubles névrotiques, troubles liés à des facteurs de stress, et troubles somatoformes	Phobies sociales, trouble obsessionnel-compulsif, réaction aigüe à un facteur de stress
F5	Syndromes comportementaux associés à des perturbations physiologiques et à des facteurs physiques	Anorexie mentale, terreurs nocturnes
F6	Troubles de la personnalité et du comportement chez l'adulte	Personnalité paranoïaque (à différencier d'une schizophrénie paranoïde), personnalité anxieuse
F7	Retard mental	
F8	Troubles du développement psychologique	Troubles du développement de la parole et du langage, trouble de la lecture, autisme infantile
F9	Troubles du comportement et troubles émotionnels apparaissant habituellement durant l'enfance ou à l'adolescence	Troubles hyperkinétiques, angoisse de séparation dans l'enfance,...

TABLE 6.8 – Exemples pour les codes F (troubles mentaux et troubles du comportement)

Dans cette section, nous ne ferons plus de mosaic plot, mais simplement des histogrammes. En effet, en observant les deux types d'illustrations pour les cas à venir, nous nous sommes rendu compte que les mosaic plot étaient moins clairs à interpréter. De plus, certains codes, tels que F8, F0 et F9, sont tellement peu présents proportionnellement aux autres que le mosaic plot possède des rectangles tellement fin que nous ne pouvons pas voir la proportion de chaque modalité de la donnée socio-démographique. Le même phénomène se produit pour la prise en charge proposée avec la rééducation qui ne concerne même pas 50 patients. Afin de ne pas tirer de conclusions hâtives sur des catégories ne comprenant que très peu d'effectifs, nous avons retiré ces modalités des graphiques. Au niveau des diagnostics, retirer ces codes semble inévitable puisque ce sont des problèmes généralement liés à l'enfance.

Remarquons que, afin d'avoir un point de comparaison et de se rappeler de la répartition des variables socio-démographiques dans l'ensemble des données, nous avons ajouté dans les histogrammes une partie "total" qui somme toutes les données utilisées dans la figure.

Sexe

Nous commençons par le sexe et le diagnostic à la FIGURE 6.17. Nous constatons que les femmes sont plus touchées par les codes F5, F3, F4 et Z, alors que les hommes le sont plutôt par F1, F6, F2 et F7. Remarquons que lorsque les femmes sont majoritaires, elles le sont toujours à plus de 60%. Par conséquent, même s'il y a 56% de femmes dans l'ensemble des données, cela est significatif. Les autres codes, donc tous les diagnostics médicaux, concernent davantage les hommes. Globalement, les femmes semblent plus touchées par le stress(F4), les facteurs influant sur les états de santé (Z), les troubles de l'humeur (F3) et les syndromes comportementaux associés à des perturbations physiologiques et à des facteurs physiques (F5), et les hommes par les autres troubles.

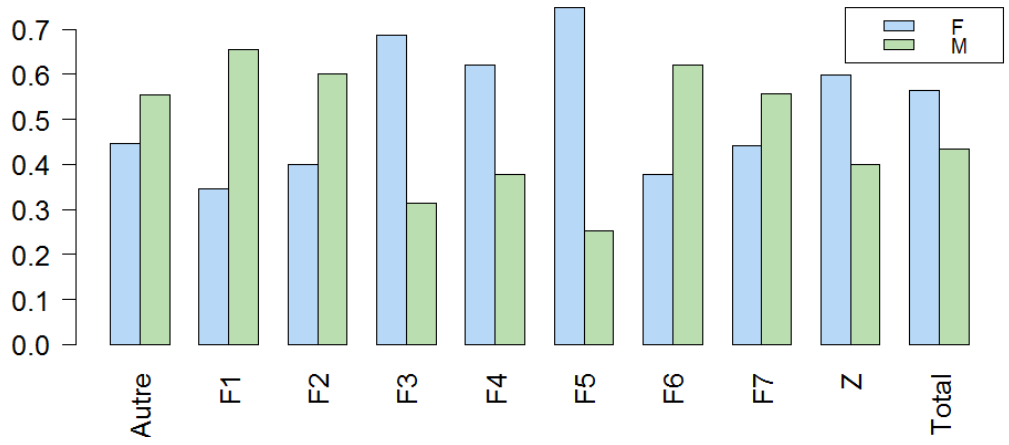


FIGURE 6.17 – Histogramme du sexe en fonction du code principal

Pour les prises en charge proposées, la répartition au sein de chaque sexe est à la FIGURE 6.18. Les hommes ont plus souvent comme prise en charge proposée un bilan/une expertise, une autre sorte ou un traitement médical. Cela est clairement significatif puisque dans ces catégories nous retrouvons une majorité d'hommes, alors que l'ensemble des données comprend presque 56% de femmes. Les femmes sont plus concernées par la proposition d'une thérapie, d'un accompagnement ou un soutien, d'un accompagnement social ou d'une information. Nous observons que les hommes ont plus de propositions réellement concrètes, une prise de médicament ou un bilan, alors que les femmes sont plus dans le soutien moral et l'accompagnement.

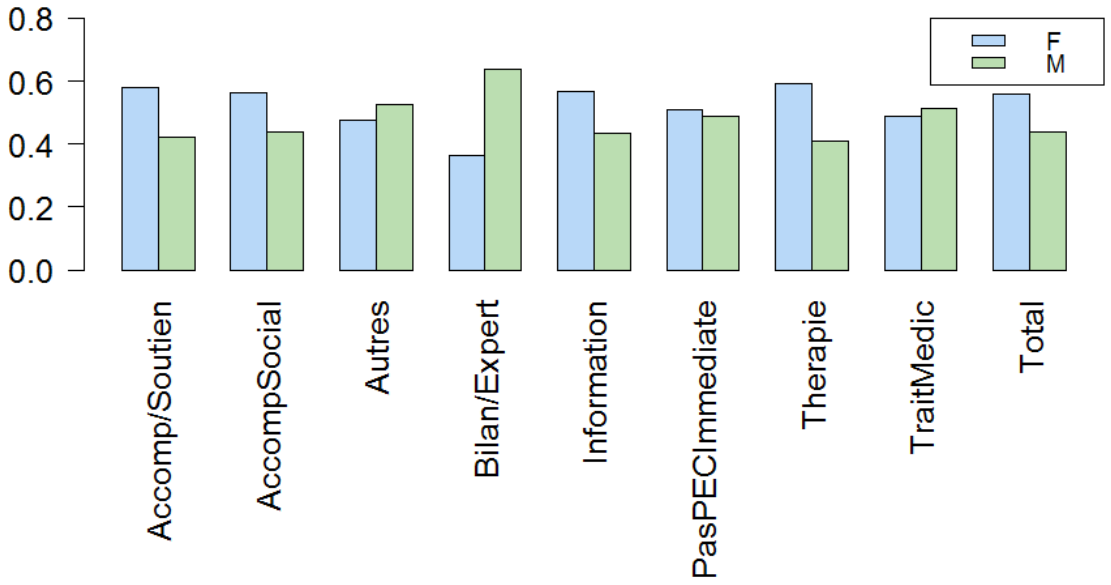


FIGURE 6.18 – Histogramme du sexe en fonction de la prise en charge proposée

Âge

Passons à la classe d'âge. La FIGURE 6.19 nous montre la correspondance de cette donnée avec les codes ICD et nous allons l'interpréter.

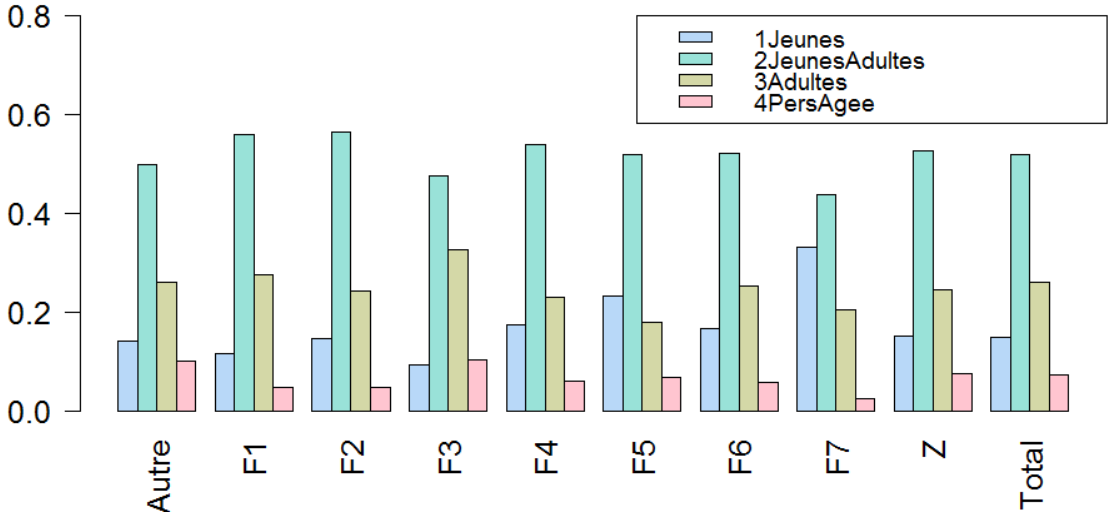


FIGURE 6.19 – Histogramme de l'âge en fonction du code

Les jeunes, entre 18 et 25 ans, semblent plus représentés dans les codes F7 et F5, donc, respectivement les "retards mentaux" et "les syndromes comportementaux associés à des perturbations physiologiques et à des facteurs physiques, tels que anorexie mentale, terreurs nocturnes,..." Les jeunes adultes sont majoritaires dans tous les codes, mais on peut voir que leur proportion n'est pas toujours la même. En effet, ils sont moins dans F7 et plus dans F1 et F2. De plus, au niveau des troubles de l'humeur, F3, nous pouvons voir que nous avons moins de jeunes et de jeunes adultes par rapport aux autres catégories, et plus d'adultes et de personnes âgées. Afin de mieux voir les différences d'âges au sein de ce diagnostic par rapport à l'ensemble des données, nous pouvons observer la FIGURE 6.20. Nous constatons que le pic de la vingtaine observé dans l'ensemble des données n'est pas du tout marqué au sein des codes "F3". Les troubles de l'humeur semblent beaucoup plus concentrés entre 35 et 55 ans. Ensuite, la courbe de ce type de code décroît moins vite.

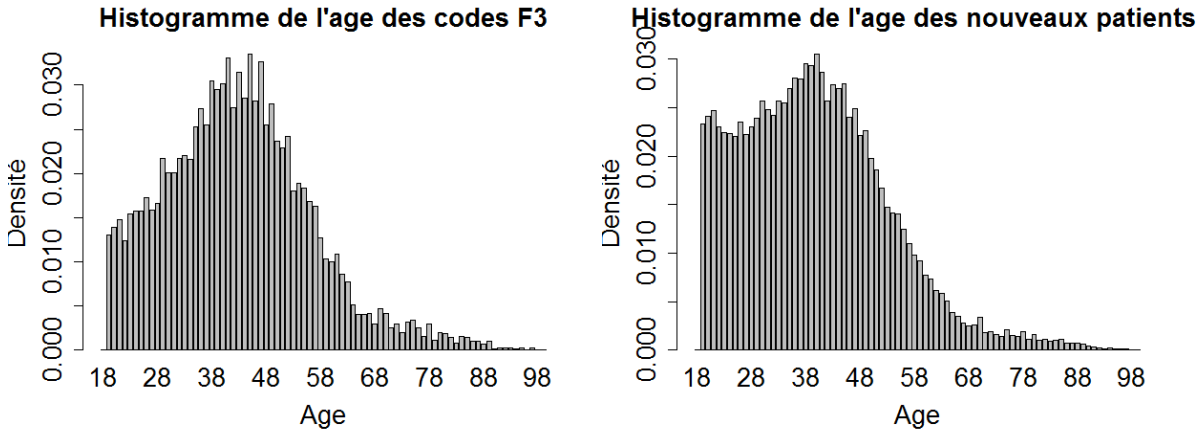


FIGURE 6.20 – Histogramme de l'âge pour les codes F3 et l'ensemble des nouveaux patients

Nous pouvons maintenant analyser les classes d'âge au sein des propositions de prise en charge, à la FIGURE 6.21. La répartition des âges semble globalement semblable au sein de chaque proposition, mis à part la modalité "bilan et expertise" qui concerne plus de jeunes et moins de jeunes adultes que les autres modalités. Cela est confirmé en faisant l'histogramme complet des âges des personnes dans cette catégorie à la FIGURE 6.22. Nous avons fait en sorte que l'échelle et les pas des deux graphiques soient identiques. Nous observons clairement un grand pic avant 25 ans et un deuxième plus petit autour de la cinquantaine. Cette FIGURE 6.22 exprime la réelle différence d'âge entre les personnes de cette catégorie et l'ensemble des données. Remarquons que l'histogramme de gauche représente un total de 2366 nouveaux patients, soit 5.6%. Cela explique le fait que la courbe soit légèrement plus saccadée.

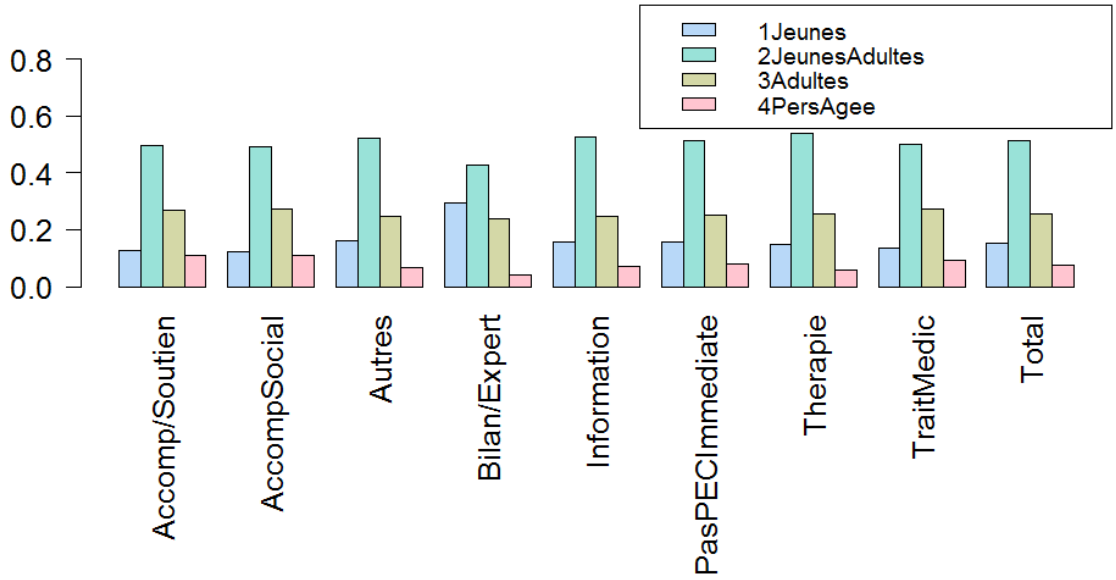


FIGURE 6.21 – Histogramme de l'âge en fonction de la prise en charge proposée

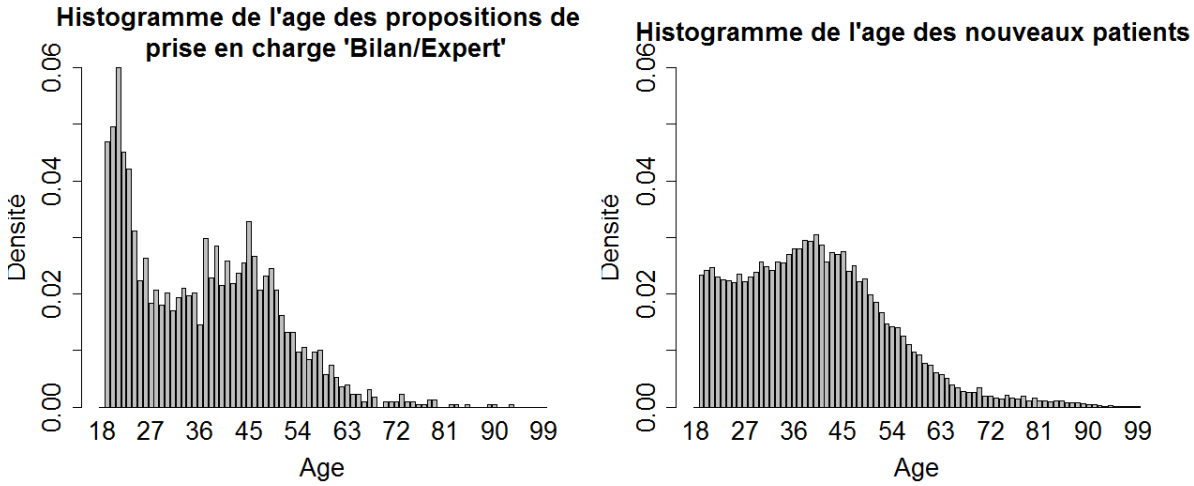


FIGURE 6.22 – Histogramme de l'âge des propositions de prise en charge "Bilan/expert"

Mode de vie

Passons au mode de vie des patients et commençons par analyser la répartition de cette donnée en fonction des codes ICD10, à la FIGURE 6.23. Le cadre familial est majoritaire dans tous les types de code, mais on constate que le "F2","schizophrénie, trouble schizotypique et troubles délirants", a une allure différente. En effet, les personnes vivant seules sont presque aussi présentes que celles en famille au sein de cette modalité. On a également une grande proportion de personnes vivant seules dont le code est "F1", donc les "Troubles mentaux et troubles du comportement liés à l'utilisation de substances

psychoactives". Au niveau des autres types de mode de vie, on peut voir qu'ils sont plus représentés dans les F2, F6 et F7, mais afin d'avoir une analyse plus précise de cette modalité, nous allons reprendre séparément les catégories incluent dans "Autre".

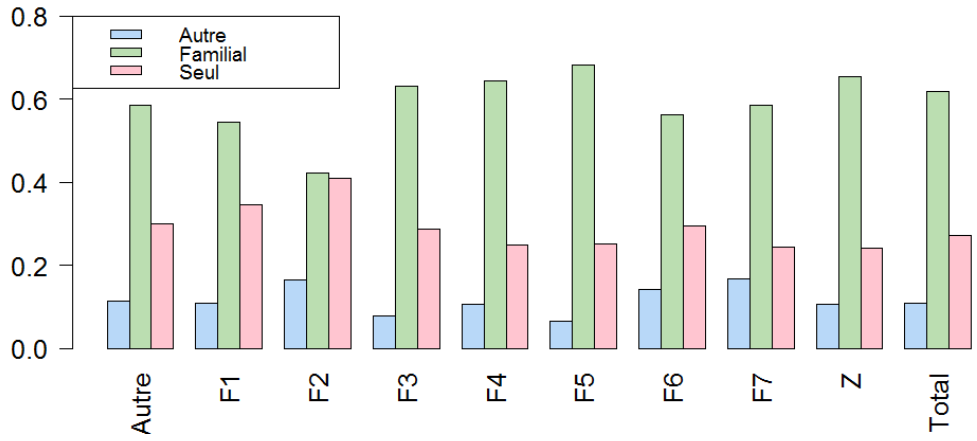


FIGURE 6.23 – Histogramme du mode de vie en fonction du code

La FIGURE 6.24 reprend les répartitions des autres modes de vie au sein de chaque code. Nous pouvons voir que le code F7, soit les retards mentaux, concerne plutôt les patients en institution. Ensuite, F4, les troubles névrotiques, comporte une majorité de personnes en milieu communautaire et F1 et F6 de patients en prison ou défense sociale. Les codes Z semblent répartis uniformément au sein des 4 autres modes de vie, alors que le total n'est pas uniforme. On remarque également qu'il y a autant de patients de milieu communautaire et d'institution parmi les codes F5, "syndromes comportementaux associés à des perturbations physiologiques et à des facteurs physiques".

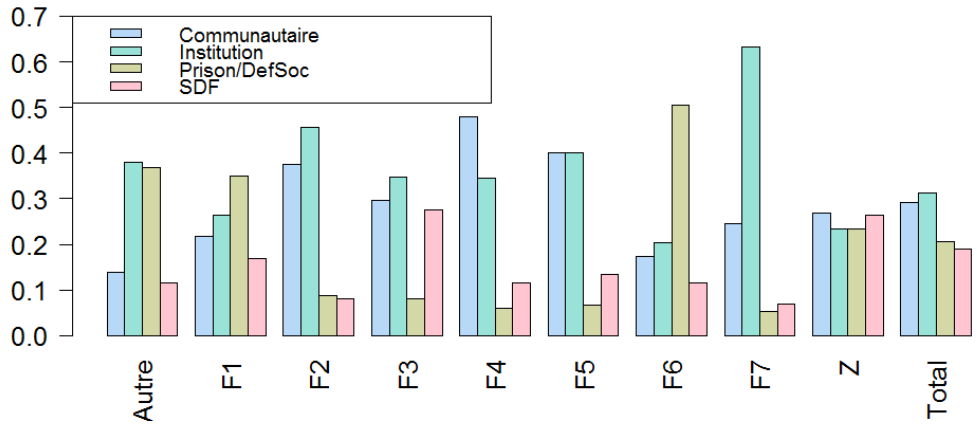


FIGURE 6.24 – Histogramme du mode de vie "Autre" en fonction du code

Pour terminer, représentons le mode de vie pour les prises en charge proposées à la FIGURE 6.25. Cet histogramme indique des différences moins flagrantes que pour certaines autres variables. Malgré tout, l'accompagnement social a été proposé à plus de personnes seules et moins de patients de milieux familiaux. En effet, il suffit de comparer la tendance de ces bâtons avec ceux du total pour observer cela. Ensuite, la thérapie semble plus suggérée aux patients vivant en famille.

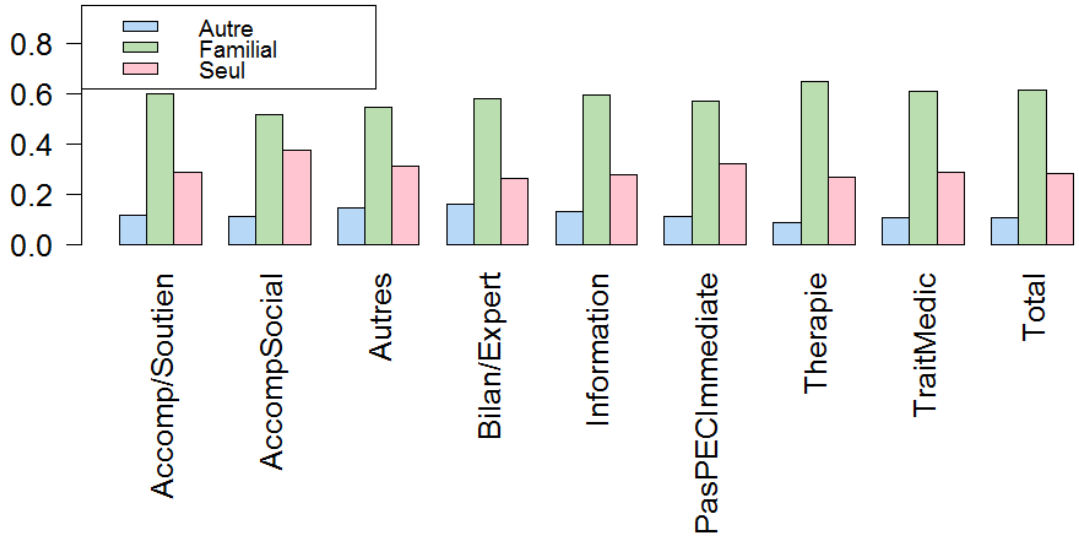


FIGURE 6.25 – Histogramme du mode de vie en fonction de la prise en charge proposée

Afin de mieux analyser les autres modes de vie, observons la FIGURE 6.26. Nous pouvons conclure de cette figure que la répartition des modes de vie au sein de chaque prise en charge proposée est différente. En effet, nous n'avons quasiment aucun patient en prison qui s'est vu proposé un traitement médical en prise en charge principale. De plus, pour les personnes en institution, nous avons plus d'accompagnements sociaux, de bilans et expertises et de traitements médicaux. Les personnes des milieux communautaires ont plus de thérapies proposées en première prise en charge. L'accompagnement social est également beaucoup proposé au SDF.

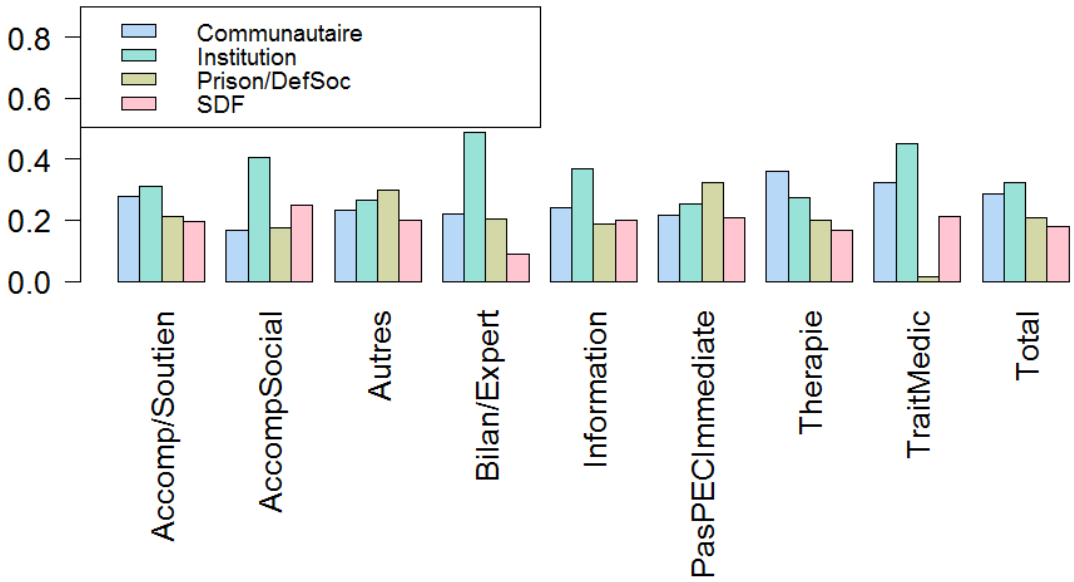


FIGURE 6.26 – Histogramme du mode de vie "Autres" en fonction de la prise en charge proposée

Nous pouvons conclure de ce chapitre que beaucoup de liens existent entre les différentes données socio-démographiques et les paramètres de la consultation. Rappelons que nous ne pouvons conclure aucun lien de causalité avec les analyses développées ici. En effet, le mode de vie pourrait influencer le problème du patient, mais son problème peut également le pousser à vivre dans un certain mode de vie. De plus, il faut bien se rappeler que nous n'avons ici que les nouveaux patients, et que donc cette étude analyse uniquement les raisons d'une première venue. C'est important lors des analyses d'âge, car nous ne pouvons pas affirmer que tel ou tel diagnostic touche les jeunes en général, mais simplement que, pour une première venue, nous avons plus de jeunes avec ce problème.

Chapitre 7

Classification

Le dernier chapitre de ce mémoire consiste à réaliser une classification des personnes présentes dans les données. De cette manière, il sera peut-être possible d'identifier des groupes de nouveaux patients ayant des caractéristiques similaires pour certaines variables.

Une classification consiste à trouver des groupes d'individus, de sorte à ce que les individus au sein d'une même classe soit les plus semblables possible et ceux dans des classes différentes les plus différents possible. Pour pouvoir quantifier la notion d'"individus proches", nous avons besoin d'une distance entre les patients et d'un indice d'agrégation. La distance permet de quantifier la façon dont les individus sont proches ou éloignés, et l'indice d'agrégation est la méthode choisie pour calculer la distance entre deux classes (composées d'individus). Afin de pouvoir utiliser les algorithmes de classification codés dans les différents programmes, tels que SAS ou R, nous devons donc avoir une notion de distance entre deux personnes. Cette mesure doit être déterminée par chacune des caractéristiques des deux individus dont on calcule la distance. Deux personnes ayant répondu la même chose doivent se trouver à une distance nulle. Plusieurs indices d'agrégation sont programmés dans les logiciels. Il nous suffira de choisir et comparer les résultats.

Dans notre base de données, toutes les variables sont de type "qualitatives non ordonnées". Cela signifie que ce ne sont pas des nombres, mais des modalités, et que celles-ci ne sont pas ordonnées. En effet, il n'y a pas de raison de considérer, par exemple, une langue maternelle avant l'autre. Par conséquent, nous ne pouvons pas simplement remplacer chaque modalité par un chiffre afin de pouvoir calculer une distance euclidienne. En effet, si par exemple, nous faisions les remplacements de la Table 7.1, nous aurions dans la classification que le Français est plus éloigné de l'Anglais que de l'Allemand. Cette méthode n'est possible que lorsque les variables qualitatives sont ordonnées.

LangueMaternelle	Code associé
Francais	1
Neerlandais	2
Allemand	3
Anglais	4

TABLE 7.1 – Exemple expliquant l'intéret de la MCA

Nous avons cherché dans la littérature ce qui avait déjà été développé afin de classifier des données qualitatives. Une première approche proposée par les aides des logiciels SAS et R consiste à procéder à une étape préliminaire calculant d'abord une analyse de correspondances multiples (MCA) et enregistrant, pour chaque personne, ses coordonnées dans ce nouvel espace défini dans un hypercube. Nous comparons le résultat à un hypercube, car les directions des nouvelles variables doivent nécessairement être perpendiculaires dans cette approche. Une fois que cette MCA sera faite, nous aurons des données quantitatives sur lesquelles la norme euclidienne pourra être appliquée. Par conséquent, nous pourrons commencer la classification.

Dans un second temps, nous avons décidé d'essayer de créer une matrice de distance entre les individus de manière intuitive. En effet, nous nous sommes dit que la distance entre deux individus pouvait être le nombre de données pour lesquelles ils n'ont pas la même réponse. Par exemple, un jeune étudiant aura une distance de 1 avec un jeune employé et de 2 avec un adulte indépendant. Une fois cette matrice créée, nous pouvons procéder à une analyse de classification. De manière plus formelle, nous pouvons définir notre norme, entre l'individu i et j, de la manière suivante :

$$d(i,j) = |var| - \sum_{k=1}^{|var|} \delta_{i_k j_k}$$

Avec $|var|$ le nombre de variables, i_k la modalité de l'individu i dans la variable k ,j_k la modalité de l'individu j dans la variable k et δ_{ij} le symbole de Kronecker, qui vaut 1 si $i = j$ et 0 sinon. Nous comptons donc bien le nombre total de variables pour lesquels les réponses pourraient être différentes et on retire le nombre de réponses égales.

Nous pouvons mentionner qu'il existe deux grands types de classification. Le premier est la classification ascendante hiérarchique, qui consiste à considérer dans un premier temps tous les individus séparés et de mettre ensemble ceux qui se ressemblent le plus. En continuant de la sorte, on groupe de plus en plus de personnes, on assemble des classes et à la fin, une seule classe reprend tous les patients. La deuxième façon de procéder à une classification est la méthode de partitionnement, qui démarre avec toutes les personnes dans la même classe et divise une classe en deux à chaque étape. En continuant jusqu'au bout, on obtient tous les individus séparés.

Pour savoir où arrêter ces algorithmes, il faut savoir combien de classes différentes sont pertinentes. Pour ce faire, nous analyserons l'ensemble des classes créées à chaque étape, avec la valeur de l'indice d'agrégation et nous déciderons du moment où assembler ou diviser des classes n'est plus efficace.

Remarquons que pour éviter d'avoir trop de données manquantes au sein des variables utilisées pour la classification, nous n'avons pris que le choix principal lorsque plusieurs réponses étaient possibles. De plus, le but de cette classification étant de déterminer des classes pertinentes au niveau socio-démographique et pathologique, nous avons gardé uniquement les variables pertinentes : TypeDossier, Sexe, Nationalite, ModedevieNiveau, CategorieProfessionnelle, SourceRevenu, NatureDemarche, DemandeduConsultant, MotifsNiveau, MotifsPremiereConsultation, CodePrincClasse, PropositiondePriseenCharge1Niveau, Ressources-principales, Province et ClasseAge. De plus, cela permet de réduire la complexité du calcul et d'obtenir des résultats en un temps raisonnable.

Nous allons faire la classification avec deux bases de données différentes. Une première ne comprenant que les variables concernant le patient, c'est-à-dire celles mentionnées ci-dessus sans le code et la proposition de prise en charge. Le niveau du motif sera également retiré dans cette partie, afin de ne garder que les motifs spécifiques. Dans un second temps, nous ajouterons ces données retirées afin de voir si la classification est différente ou si ces variables n'ont pas d'impact.

Remarquons que dans cette partie, les "DM" sont considérées, dans un premier temps, comme une modalité. En effet, cela permet de conserver tous les patients. Pour que la procédure de classification puisse opérer, il faut réduire la taille de la base de données. Nous allons donc, lorsque cela s'avère nécessaire, procéder à un échantillonnage aléatoire de 3875 patients, comme vu précédemment. Prendre plus de patients générait des matrices trop volumineuses qui inséraient des erreurs dans R. Nous avons procéder plusieurs aux analyses avec 3875 patients pour nous assurer que nous n'avions pas un échantillon trop éloigné de l'ensemble des données. Dans un second temps, pour avoir une classification sans données manquantes, nous retirerons toutes les personnes dont au moins une donnée manque. Après cette opération, la première base de données considérée comprendra 22850 patients, et la deuxième 20820.

Ce chapitre va d'abord analyser la première base de données, puis la deuxième et enfin, conclure à une classification. Dans chaque section, nous allons identifier la matrice de distance utilisée, déterminer le nombre de classes, procéder aux classifications et interpréter les résultats.

7.1 Variables du patient

Une première classification, reprenant uniquement les variables du patient et conservant les données manquantes comme une modalité, est effectuée dans cette section.

7.1.1 MCA

Création d'une matrice de distance

Nous commençons par associer des valeurs numériques aux individus de l'ensemble des données à l'aide d'une analyse de correspondances multiples (MCA). Nous obtenons un total de 80 valeurs propres, dont la fonction cumulative de la variance expliquée correspondante se trouve à la FIGURE 7.1. Ce graphique nous indique qu'il n'y a pas de direction que l'on peut retirer tout en gardant malgré tout la majorité de la variance expliquée. En effet, la courbe augmente sans avoir de réel changement de pente. De plus, dans notre cas, même si nous décidons de garder les 80 valeurs propres, cette méthode nous permet de transformer chaque individu en un vecteur numérique de 91 dimensions.

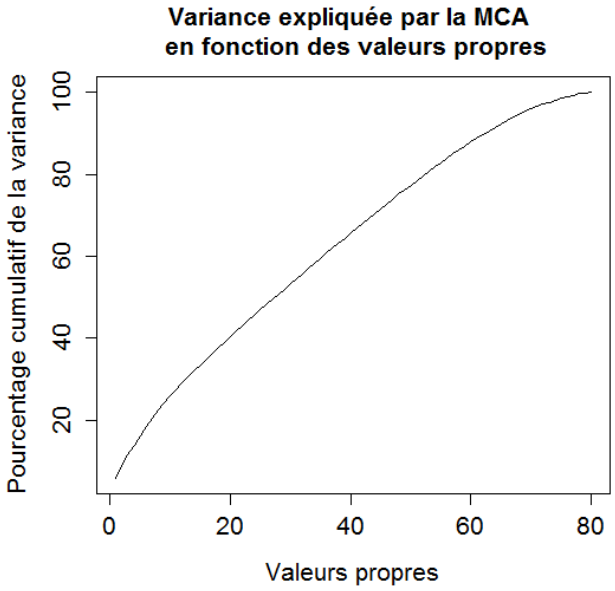


FIGURE 7.1 – Analyse des correspondances multiples

Expliquons la raison des 80 valeurs propres et des 91 dimensions. Nous avons dans la base de données considéré 12 variables différentes. Chaque variable comprend plusieurs modalités possible. Par exemple, la donnée "sexe" possède 3 modalités : hommes, femmes et "DM". Si nous comptabilisons le nombre total de modalités dans les données, nous en obtenons 93. Donc notre hypercube démarre avec 93 dimensions. Certaines modalités sont tellement sous-représentées qu'elles n'apparaissent plus dans la MCA. Nous avons donc maintenant 91 dimensions. De plus, remarquons que parmi les 93 modalités de départ, certaines peuvent être déduites par les autres. En effet, dans notre exemple,

nous pouvons par exemple retirer la modalité "femmes", car si un patient n'a ni la modalité "hommes", ni "DM", nous pourrons en conclure que c'est une femmes. Par conséquent, dans chaque variable, une dimension dépend des autres et est simplement la direction opposée à toutes les autres modalités de la variables. Il nous reste donc $93 - 12 = 81$ modalités indépendantes. Par conséquent, le nombre de degré de liberté, donc de valeurs propres, est $81 - 1 = 80$.

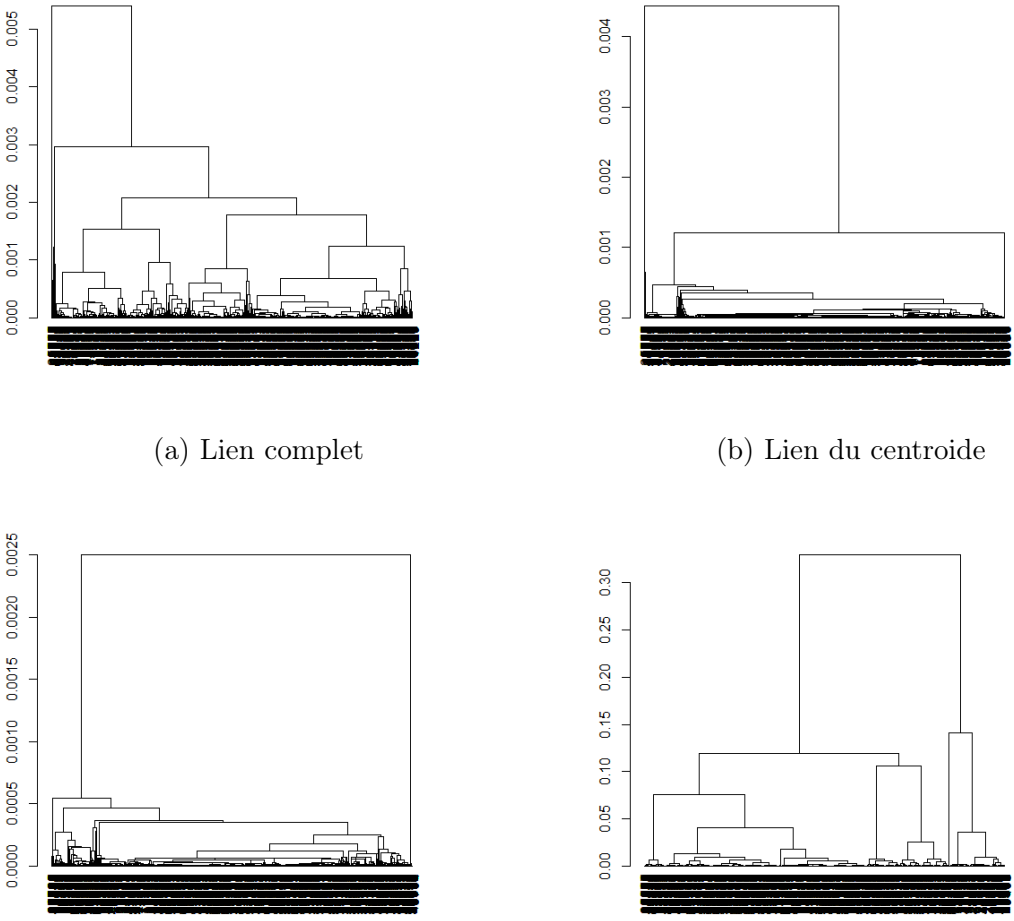
Nous constatons que, par la suite, nous ne parviendrons pas à représenter significativement les données en 2 dimensions. Cependant, grâce à la MCA, nous créons une nouvelle base de données où chaque individu est décrit par 91 valeurs numériques au lieu de 12 modalités. Maintenant que les données sont de type quantitatif, l'algorithme de classification pourra générer une matrice de distance à l'aide de la norme euclidienne. Nous pouvons donc passer à la détermination du nombre de classes.

Détermination du nombre de classes

Nous allons procéder à une classification ascendante hiérarchique. Dans cette partie, garder les 41916 individus empêchait R d'arriver au bout de la classification. Nous avons donc pris un échantillon aléatoire de 3875 patients. Puisque cette méthode part de chaque individu séparément et fait des classes au fur et à mesure, nous pouvons la représenter à l'aide d'un dendrogramme. Ce genre de représentation contient tous les individus en bas du graphique, et des "ponts" permettent de montrer lesquels sont joints à quel moment. La hauteur de ce lien est proportionnelle à la valeur de l'indice d'agrégation. Par conséquent, lorsque le saut est haut, cela signifie que la méthode a rassemblé deux classes plutôt éloignées. Pour choisir le nombre de classes, nous arrêterons donc l'algorithme lorsqu'il y aura un palier assez haut, car cela signifie que si on joint ces deux classes, des éléments assez éloignés seraient ensemble.

Le choix de la manière dont les classes vont être formées, donc de l'indice d'agrégation, va influencer le dendrogramme. Nous avons fait l'analyse pour plusieurs indices d'agrégation et les dendrogrammes correspondants se trouvent à la FIGURE 7.2. Nous avons le lien complet (FIGURE 7.2a), le centroïde (FIGURE 7.2b), la médiane (FIGURE 7.2c) et Ward (FIGURE 7.2d).

Nous constatons que la plupart des plus grands sauts se trouvent lorsqu'on rassemble les deux dernières classes en une. Cependant, on peut voir que pour tous les indices d'agrégation sauf Ward, la classification en deux parties est composée d'une énorme classe et d'une autre ne comprenant que quelques individus. Pour le lien simple et le centroide, la partition en trois classes semble également intéressante, mais contient également un grand groupe et quelques isolés. L'indice d'agrégation le plus souvent utilisé est l'indice de Ward et semble dans notre cas le plus intéressant. On y voit que les partitions en deux et six classes semblent pertinentes. De plus, celles-ci ne sont pas simplement des classes qui ne comportent que quelques patients et une grosse avec tous les autres.



(a) Lien complet

(b) Lien du centroide

(c) Lien de la médiane

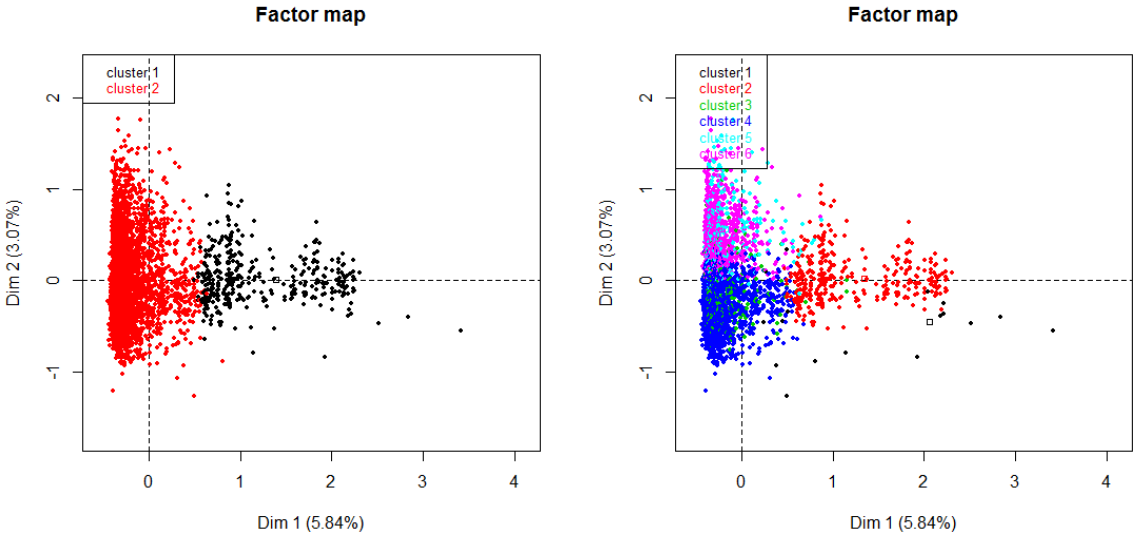
(d) Lien "Ward"

FIGURE 7.2 – Dendrogramme basé sur l'analyse des correspondances multiples

Nous allons visualiser, dans le plan des deux premières composantes principales de la MCA, les classifications en 2 et 6 parties créées avec l'indice de Ward dans la section suivante.

Classifications

La FIGURE 7.3 reprend les partitions en 2 et 6 classes représentées dans les deux premières dimensions de l'analyse des correspondances multiples. On peut voir que la classification en deux classes sépare clairement en fonction de la première dimension. On a une classe avec les grandes valeurs et une autre avec les plus petites. Pour ce qui est de la partition en 6 classes, nous avons une classe, en rouge, avec les grandes valeurs dans la dimension 1 ; une autre en rose avec les petits dim1 et les valeurs positives de la deuxième dimension ; puis une en bleu avec les petits dim1 et les dim2 négatifs. Les trois autres classes ne sont pas clairement distinguables dans ce plan.



(a) 2 classes (b) 6 classes

FIGURE 7.3 – Partition en 2 et 6 classes via la MCA

Le plan des dimensions deux et trois permet de percevoir la classe 5, en bleu clair, à la FIGURE 7.4. Cette classe identifie les valeurs négatives de la dimension 3. Pour ce qui est des interprétations en fonction des modalités des variables initiales, elles seront faites plus tard, lorsque toutes les classifications auront été définies.

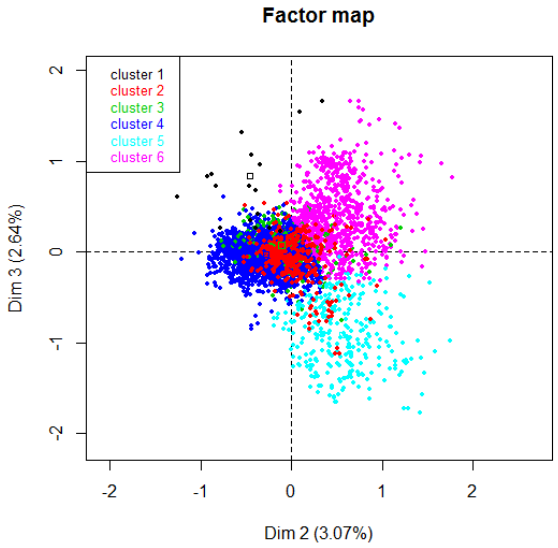


FIGURE 7.4 – MCA - 6 classes dans dimensions 2 et 3

7.1.2 Intuitivement

Détermination de la matrice de distance

Nous avons décidé de suivre une approche plus intuitive, afin de voir où cela nous mène. En effet, afin de classer les personnes, il faut pouvoir quantifier la distance entre deux individus différents. Le passage par une MCA est appelé "tandem analysis" [27], car le procédé est ainsi séparé en deux parties distinctes dont le but n'est pas commun. L'utilisation de la MCA pourrait retirer de l'information importante pour la classification.

Nous avons dans un premier temps essayé de créer la matrice de distance intuitive pour l'ensemble des données, mais cela engendrait une matrice trop grande, que le logiciel R ne pouvait pas gérer. Nous avons donc décidé de faire la classification sur un échantillon aléatoire. Comme décrit précédemment, nous choisissons un échantillon de 3875 personnes choisies de manière aléatoire avec la même probabilité pour chaque nouveau patient d'apparaître dans l'échantillon.

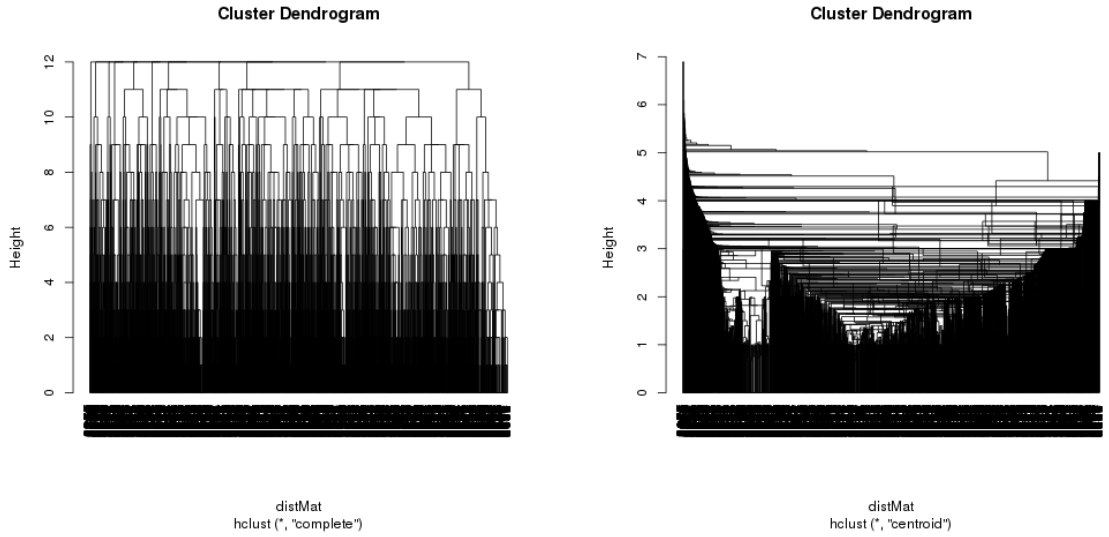
On a donc créé, à l'aide de R, une matrice D de dimension 3875×3875 dont chaque élément D_{ij} correspond au nombre de réponses différentes des individus i et j. Cette matrice est donc symétrique et sa diagonale est nulle. Pour pouvoir illustrer les classes par la suite, nous conservons en mémoire les index des individus choisis lors de l'échantillonnage.

Maintenant que cette matrice de distance est déterminée, nous allons pouvoir procéder à la classification. Pour ce faire, nous allons utiliser le logiciel R pour calculer une classification ascendante hiérarchique. Nous nous baserons sur les diapositives de A. Bouchier [4] pour faire cette section.

Détermination du nombre de classes

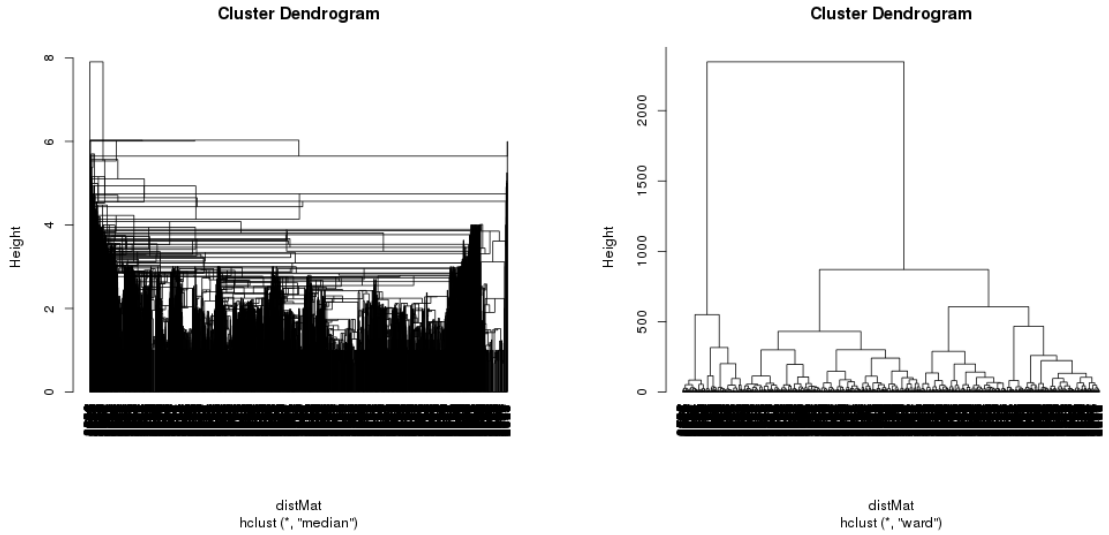
Ayant appliqué une méthode ascendante hiérarchique, nous pouvons représenter les différentes étapes de la classification à l'aide d'un dendrogramme. Nous pouvons choisir plusieurs types d'indices d'agrégation différents. Certains indices d'agrégation sont plus efficaces dans certains cas. Nous pouvons voir, à la FIGURE 7.5, les différents dendrogrammes pour le lien complet (7.5a), le centroide (7.5b), la médiane (7.5c) et Ward (7.5d).

Nous constatons que dans notre cas, la plupart des dendrogrammes ne peuvent pas être interprétés correctement, car on ne distingue aucun palier. Par contre, la méthode de Ward, qui est la plus souvent utilisée, nous permet d'avoir un dendrogramme clair. On peut y voir un énorme saut lorsque l'on rassemble les deux dernières classes, nous indiquant que les deux classes que l'on rassemble à cette étape étaient assez éloignées. Il faut donc garder ces deux classes distinctes. Nous pouvons voir que couper à trois classes paraît également envisageable, car ce palier est bien distinct des autres. On envisagera donc une partition en 2 ou 3 classes.



(a) Lien complet

(b) Lien du centroide



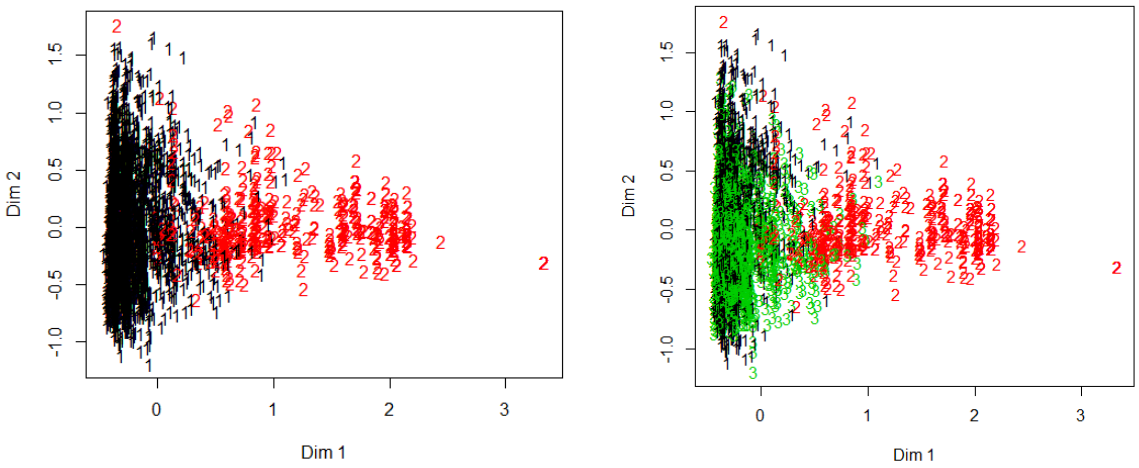
(c) Lien de la médiane

(d) Lien "Ward"

FIGURE 7.5 – Dendrogramme correspondant à la distance intuitive

Classifications

Maintenant que nous avons choisi d'analyser les partitions en 2 et 3 classes, créons-les et essayons de les visualiser. Puisque l'indice d'agrégation retenu précédemment était celui de Ward, nous allons continuer avec cette méthode. Dans un premier temps, nous allons faire une MCA afin de visualiser où se trouve chaque classe par rapport aux deux premières composantes principales. En procédant à cette MCA, nous avons une première dimension expliquant 6.20% de la variance, et une deuxième en expliquant 3.13%. Les classes obtenues représentées dans ces deux dimensions se trouvent à la FIGURE 7.6.



(a) 2 classes (b) 3classes

FIGURE 7.6 – Partition en 2 et 3 classes via la distance intuitive

On peut voir à la partition en 2 classes (7.6a) que les deux parties semblent séparer les individus avec une grande valeur dans la première dimension de ceux avec une plus petite valeur. Pour la partition en 3 classes, la classe 1 est séparée en deux. Nous pouvons voir à l'illustration 7.6b que cette subdivision ne coïncide pas avec des valeurs spécifiques d'une des deux premières dimensions de la MCA. Malgré diverses tentatives pour trouver un plan en deux dimensions de la MCA permettant d'illustrer la division de la classe 1, nous n'avons pas trouvé de graphique pertinent. Cela signifie que, soit il faut plus de deux dimensions simultanément pour que cette division ait du sens, soit la MCA ne permet pas d'identifier ces classes.

7.1.3 Interprétations

Les deux méthodes employées semblent créer deux classes. Sur les deux premières dimensions des MCA, les partitions en deux classes semblent similaires. Nous allons chercher à identifier les types de patients dans chaque classe. Remarquons que la classification intuitive sépare en deux groupes : un de 3312 éléments et un de 563. Nous avons donc une classe assez importante et une possédant moins de personnes, mais malgré tout un certain nombre. La méthode de MCA a une forme similaire, car il y a une classe de 3359 patients et l'autre de 516.

En analysant les carrés des cosinus et les coordonnées de chaque modalité au sein de la première dimension, on se rend compte que les modalités qui pèsent principalement sur cette dimension sont les données manquantes. Cependant, il y a tellement de modalités différentes que l'analyse par les composantes principales ne serait pas claire, d'autant plus qu'on sait que ces composantes nous font perdre certaines informations.

Afin de comprendre les deux partitions en deux classes, nous avons regardé la répartition de chaque variable au sein de chaque classe. Nous présentons ci-dessous ces proportions pour les données présentant des différences en fonction de la classe.

Nous avons considéré que chaque classe possédait 100% d'effectifs et observé les proportions de chaque modalité dans la classe. Par conséquent, les histogrammes sont générés de sorte à ce que la somme des modalités dans une même classe fasse 1. De cette manière, la représentation ne sera pas influencée par la différence de taille des classes. Nous avons mis côte à côte les répartitions de chaque classe et nous aurons toujours sur une même figure le graphique des classes avec la méthode de la MCA et celle intuitive.

Nous pouvons commencer par les données socio-démographiques. La répartition des provinces, sans celles sous-représentées, se trouve à la FIGURE 7.7. Nous pouvons voir que dans les deux méthodes, la classe 1 contient proportionnellement plus de personnes de Liège, du Luxembourg et de Namur, alors que la classe 2 ans a plus du Hainaut, du Brabant Wallon et de Bruxelles-Capitale. Remarquons que ce dernier est également très peu représenté et aurait pu être retiré du graphique. La tendance des deux méthodes est tout-à-fait similaire, même si de légères différences au niveau des proportions sont présentes.

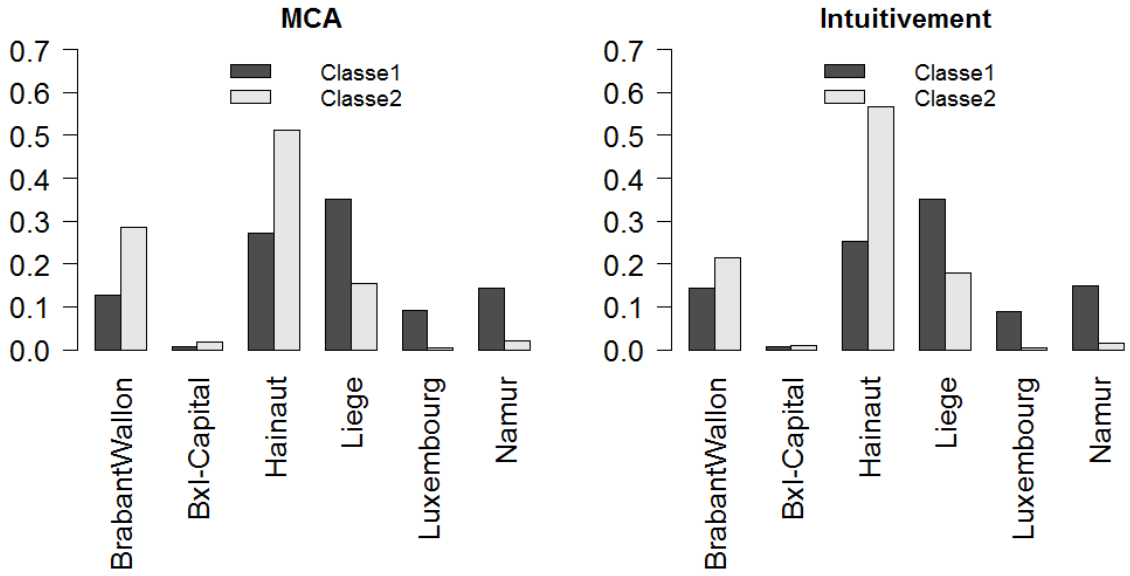


FIGURE 7.7 – Province

Notons que nous avons fait ces classifications sur un échantillon qui n'est pas le même pour chaque méthode. En effet, dans les deux cas, le code commençait par générer un échantillon, puis l'analysait. Nous avons donc exécuté plusieurs fois ce code pour nous assurer que les résultats étaient semblables et que l'échantillon ne biaisait pas les classifications.

La FIGURE 7.8 illustre les répartitions pour le mode de vie. Nous pouvons voir que les deux méthodes ressortent des résultats semblables, mais pas tout-à-fait identiques, au niveau de cette donnée. Dans les deux cas, la classe 1 est majoritairement composée de modes de vie familiaux et la classe 2 contient plutôt des personnes vivant seules, des SDF, des autres modes de vie. La deuxième classe possède également assez bien de personnes vivant dans un cadre familial. Cependant, proportionnellement, la classe 1 possède beaucoup plus de patients avec ce mode de vie. Les catégories sous-représentées ne semblent pas avoir des tendances similaires avec les deux méthodes. De plus, les personnes vivant seules sont proportionnellement presque autant représentées dans les deux classes.

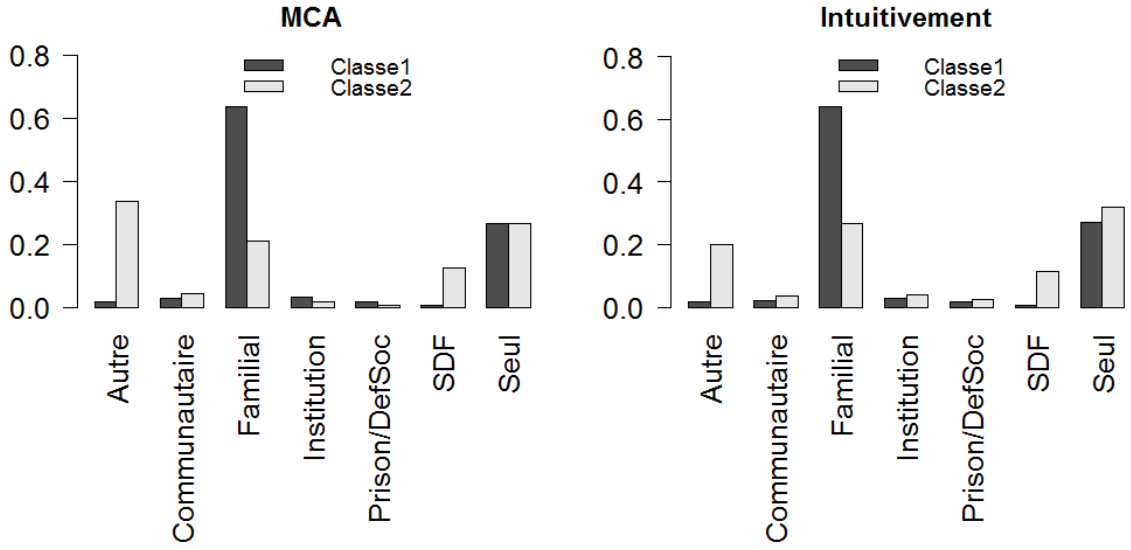


FIGURE 7.8 – Mode de vie

Passons à la catégorie professionnelle à la FIGURE 7.9. Cette variable est la première où l'on voit réellement une différence entre les deux méthodes. En effet, dans la classification intuitive, la classe 2 est composée à plus de 60% d'indépendants, à plus de 15% d'étudiants et contient la majorité des professions libérales. Les patients sans profession sont plutôt dans la classe 1 avec les employés et les ouvriers.

Par contre, la classification où nous sommes passés par une MCA contient une plus grosse proportion de personnes sans profession dans la classe 2. On a malgré tout beaucoup d'indépendants et d'étudiants, mais la proportion d'indépendants est beaucoup plus petite.

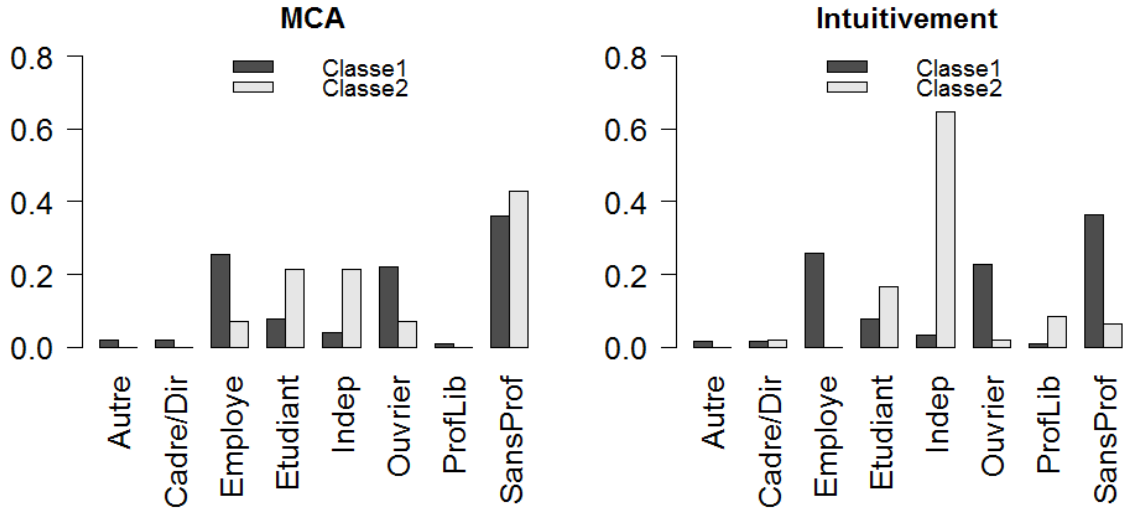


FIGURE 7.9 – Catégorie professionnelle

Passons aux sources de revenu. Nous pouvons voir à la FIGURE 7.10 que les deux méthodes aboutissent à des partitions semblables. En effet, dans les deux cas, la classe 2 est composée principalement des sources de revenu inconnues et la première classe, de tous les types de revenu connu. Cela ne nous permet pas de tirer de conclusion sur la classification de cette variable. Lorsque nous retirerons les "DM", dans la section suivante, nous retirerons également les "Inconnu".

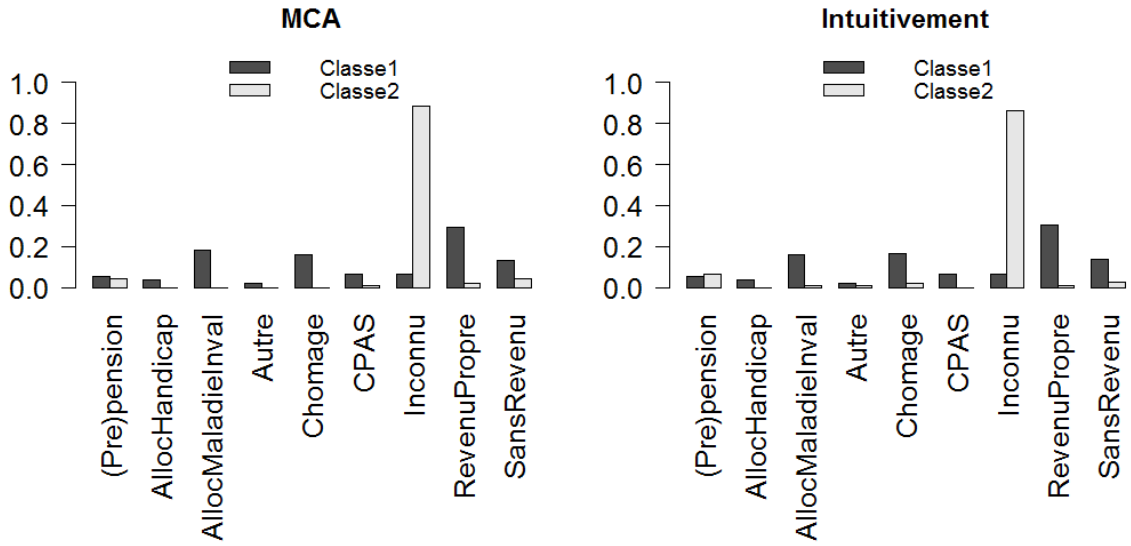


FIGURE 7.10 – Source de revenu

Nous pouvons maintenant analyser les variables concernant la consultation. Tout d'abord, la FIGURE 7.11 contient les demandes des consultants. Nous pouvons y constater que les deux classes n'ont pas de composition extrêmement différente. Cependant, avec les deux méthodes, la première classe contient légèrement plus de personnes demandant un suivi.

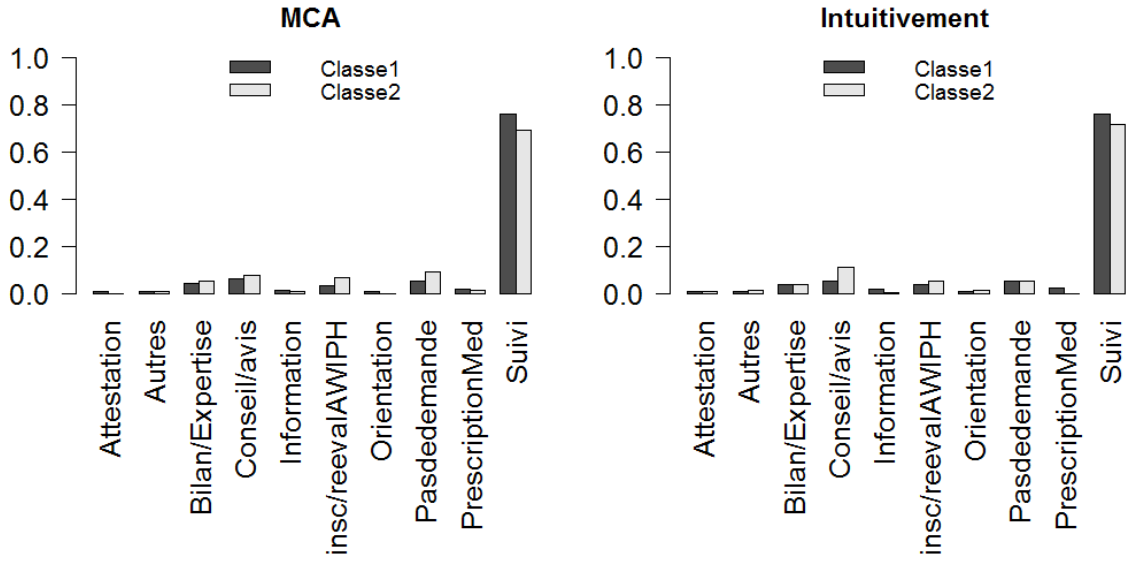


FIGURE 7.11 – Demande du consultant

La dernière donnée pertinente à analyser est la nature de la démarche, à la FIGURE 7.12. À nouveau, la méthode passant par les analyses de correspondances multiples et de la matrice de distance intuitive nous donne des répartitions similaires. On constate que la classe 1 contient plus de démarches orientées, la 2 plus de spontanées, et les proportions de démarches contraintes semblent similaires dans les deux classes.

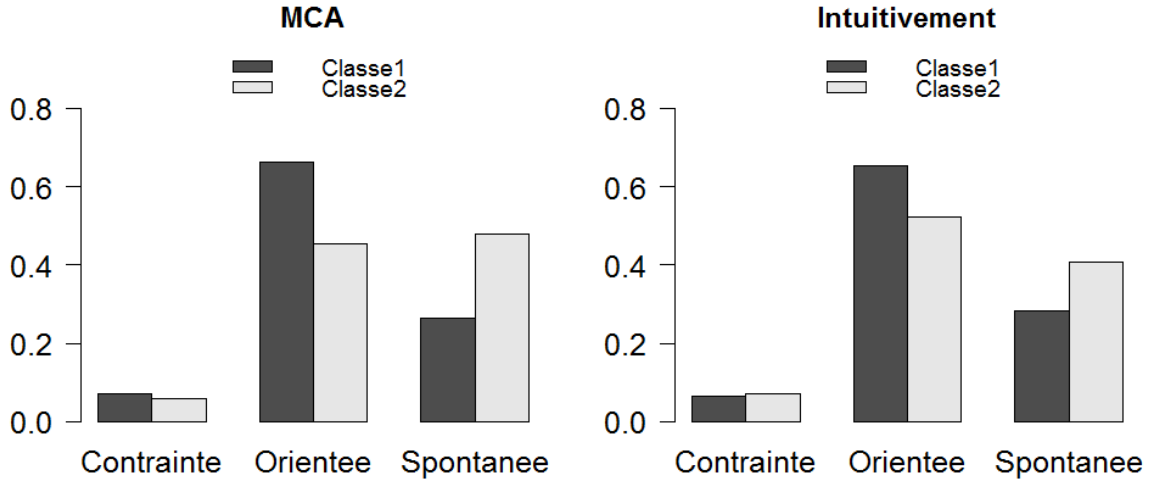


FIGURE 7.12 – Nature de la démarche

Les autres variables de cette base de données ne possédaient pas de différences significatives entre les deux classes, ni avec la MCA, ni avec la matrice intuitive. Pour rappel, ces données sont le type de dossier, le sexe, la nationalité, les motifs de la première consultation, les ressources principales et la classe d'âge.

7.1.4 Conclusion

Avec ces données, les deux méthodes donnent des résultats semblables. Nous savons que le passage par la MCA consiste en une modification des données dont le critère est simplement de diagonaliser au maximum la table de contingence. Nous pourrions donc perdre de l'information. Malgré tout, cette méthode est souvent utilisée. La matrice de distance intuitive indique le nombre de variables pour lesquelles les patients n'ont pas les mêmes réponses. Procéder à une classification basée sur ces distances a du sens, car elle va assembler les personnes qui ont la plus petite distance, donc le moins de réponses différentes. Puisque dans notre cas les deux méthodes donnent des résultats similaires, nous allons garder cette classification en deux classes. Cependant, la façon intuitive de classer les individus montraient des tendances plus prononcées.

Nous allons décrire les résultats intéressants pour les deux classes. Nous allons également indiquer les caractéristiques les plus présentes dans chaque classe. Pour les provinces, nous voyons le résultat à la FIGURE 7.13[1]. La classe en rouge correspond à la classe 1 décrite précédemment. Cette classe, reprenant plus d'habitants de Liège, Luxembourg et Namur correspond globalement aux patients vivant dans un cadre familial, étant employés, ouvriers ou sans profession et dont les démarches sont orientées. La classe bleue, reprenant le Hainaut, le Brabant Wallon et Bruxelles-Capitale correspond plutôt aux personnes vivant seules, SDF ou dans un autre mode de vie ; indépendantes, étudiantes ou dans une profession libérale et dont les démarches sont spontanées.

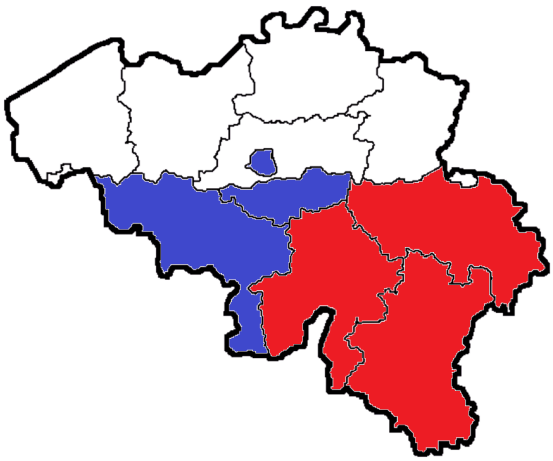


FIGURE 7.13 – Carte de la classification

1. http://www.n-joypc.be/default_fr.asp?page=20090917001655

7.2 Toutes les variables de classification

Nous allons maintenant procéder à une classification plus précise qui reprend les prises en charge proposées et les types des codes principaux. De plus, nous allons ici prendre l'échantillon aléatoire uniquement sur les patients ne possédant aucune donnée manquante dans les variables considérées. L'analyse des données manquantes faite précédemment nous permet de nous assurer que ce procédé n'est pas un problème et n'influence pas trop la classification. Comme dans la partie précédente, nous allons procéder avec une MCA, puis avec une matrice intuitive créée de manière similaire.

Dans cette partie, nous avons tiré un échantillon aléatoire que nous utilisons pour les deux méthodes. Nous pourrons donc plus aisément comparer les deux résultats.

7.2.1 MCA

Après avoir effectué une analyse de correspondances multiples, nous avons effectué plusieurs types de classifications ascendantes hiérarchiques. Le seul indice d'agrégation ressortant un dendrogramme lisible est celui de Ward. À la FIGURE 7.14, le dendrogramme associé nous indique un choix de 2, 4 ou 6 classes. En effet, nous pouvons voir un plateau plus élevé à ces niveaux, qui nous incitent à vouloir couper l'arbre à cet endroit. Nous allons observer à quoi ressemblent les partitions en 2, 4 et 6 classes dans certaines dimensions de l'analyse de correspondances multiples.

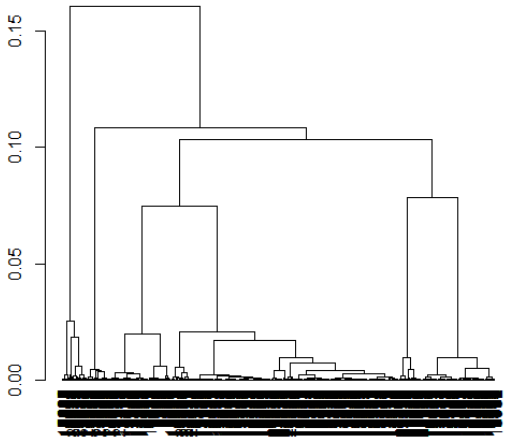


FIGURE 7.14 – Dendrogramme après MCA avec toutes les variables sans les DM

Nous pouvons commencer par observer les partitions en 2 et 4 classes dans les deux premières dimensions de la MCA. Ces deux dimensions expliquent 3.87% et 2.96% de la variance. Ce plan illustre donc 6.83% de la variance des données. À la FIGURE 7.15,

nous observons que la première classification, en deux classes, sépare les grandes et les petites valeurs de la première dimension. De plus, la partition en 4 classes a simplement redivisé la deuxième classe en 3 groupes. Le bleu correspond aux patients avec une grande deuxième dimension et la verte, ainsi que la rouge avec une petite. Pour pouvoir différencier la deuxième et la troisième classe, il faut regarder cette partition dans d'autres dimensions de la MCA, comme illustré à la FIGURE 7.16.

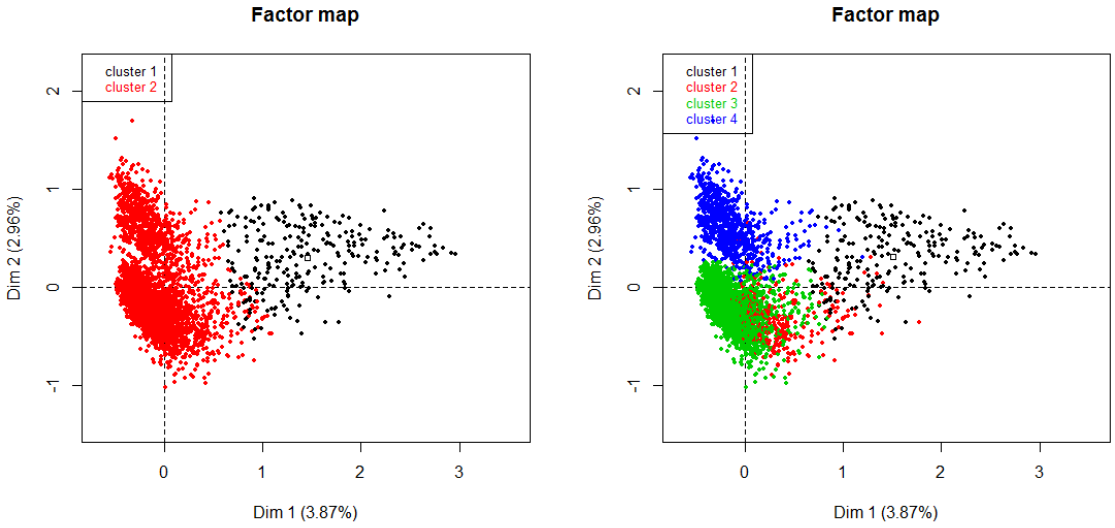


FIGURE 7.15 – 2 et 4 classes dans les dimensions 1 et 2 de la MCA.

La FIGURE 7.16 indique que les classes 2 et 3 sont en fait séparées au niveau de la troisième dimension. En effet, la 3 contient les grands indices en dimension 3 et la 2 les plus petits.

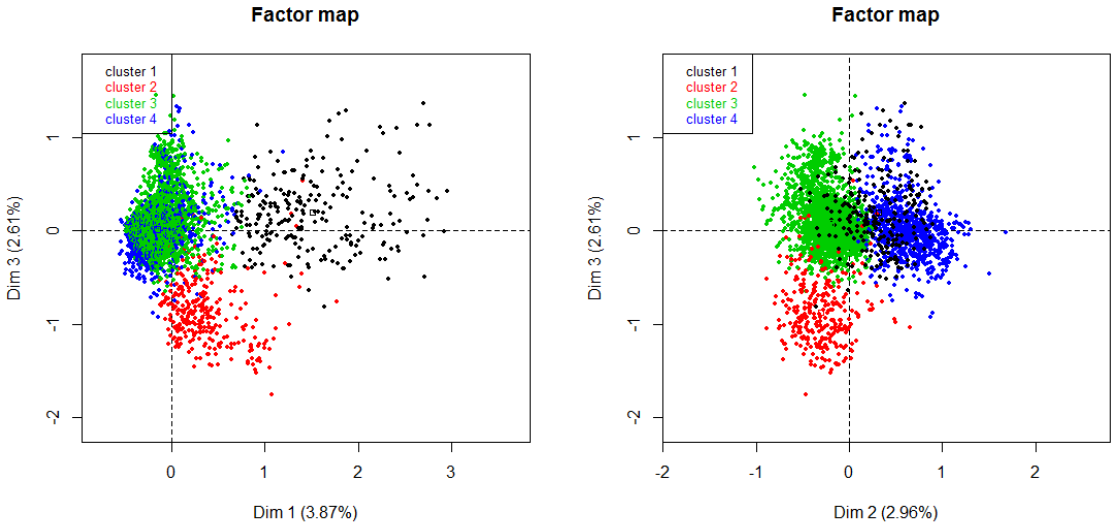


FIGURE 7.16 – 4 classes dans diverses dimensions de la MCA.

Remarquons que nous avons tellement de valeurs propres et de modalités différentes qu'il ne serait pas clair de chercher ce qui définit chaque dimension précisément. Pour malgré tout s'en faire une idée, nous allons regarder pour les trois premières dimensions, les cinq modalités avec les plus grands cosinus au carré. En effet, plus la valeur absolue du cosinus, donc son carré, est grand, plus cette modalité se trouve proche de l'axe et donc l'influence. Pour les trois premières dimensions, nous avons la Table 7.2.

dimension	Modalité	Cosinus Carré
dim 1	PECPropNiveau_Bilan/Expert	0.503
	MotifsNiveau_Autre	0.368
	DemandeduConsultant_insc/reevalAWIPH	0.356
	CodePrincClasse_F7	0.278
	DemandeduConsultant_Suivi	0.255
dim 2	MotifsNiveau_ProbPer	0.686
	MotifsNiveau_ProbRel	0.653
	Motifs_RelCouple	0.240
	Motifs_RelFamille	0.239
	CodePrincClasse_Z	0.183
dim 3	Dem_Contrainte	0.370
	Motifs_PersoActeDel	0.360
	CatProf_Etudiant	0.238
	1Jeunes	0.211
	SansRevenu	0.130

Table 7.2 – Plus grands cosinus carrés des 3 premières dimensions de la MCA

La Table 7.2 nous donne une première idée de ce qui compose les dimensions. Plus le cosinus au carré est proche de 1, plus la modalité forme un petit angle avec l'axe, c'est-à-dire en est proche. Nous pouvons donc, en quelque sorte, considérer qu'elle guide cette dimension. La première dimension est fort déterminée par les propositions de prise en charge "Bilan et Expertise", les autres motifs, les demandes d'inscription et de réévaluation AWIPH, les codes de type F7 et les demandes de suivi. Nous avons créé le tableau de sorte à avoir les modalités par ordre décroissant de cosinus carré. Les deux grands types de motifs pèsent fort sur la dimension 2. En effet, les deux plus grands cosinus au carré sont ceux correspondant aux problèmes relationnels et personnels. Ensuite, les motifs plus spécifiques ont leur importance et les codes sont de type Z. Enfin, la troisième dimension semble avoir comme extrême les démarches contraintes, les motifs personnels liés à des actes délictueux, les étudiants, les jeunes et les personnes sans revenu.

Nous constatons que les extrêmes des dimensions semblent montrer des classes de personnes assez intuitives. Par exemple, la dimension 3 concerne les jeunes étudiants qui n'ont pas encore de revenu, commettent un acte délictueux et sont contraints de consulter.

Pour pouvoir réellement justifier la nature des classes, il faudra regarder, variable par variable, la répartition de chaque modalité dans chaque classe. Nous avions également une classification en 6 classes proposée. La FIGURE 7.17 nous montre la partition obtenue dans le plan des deux premières dimensions de la MCA. À nouveau, nous pouvons voir que les classes séparent les grandes et petites valeurs de certaines dimensions.

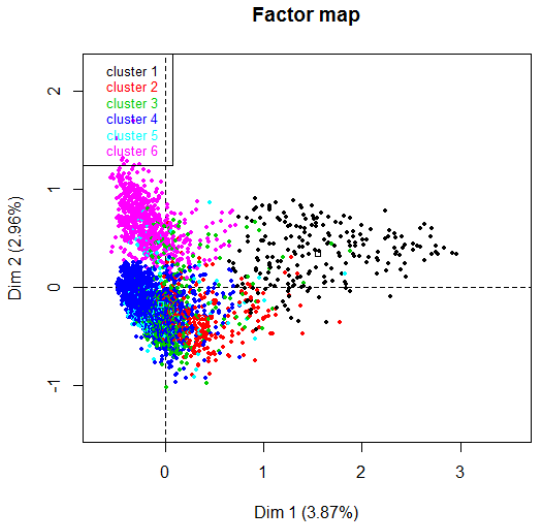


FIGURE 7.17 – 6 classes dans les dimensions 1 et 2 de la MCA.

Ceci est logique, puisque nous avons d'abord procédé à la MCA, et ensuite classé les points qui en ressortaient. Par conséquent, les points proches dans cet hypercube ont été groupés. Les classes qui nous semblent l'une sur l'autre dans ce plan nécessitent un autre angle de vue de l'hypercube de la MCA pour être différenciées. En effet, si nous regardons les dimensions 1 et 3, puis 2 et 3, à la FIGURE 7.18, nous distinguons désormais bien les classes qui nous paraissaient mélangées.

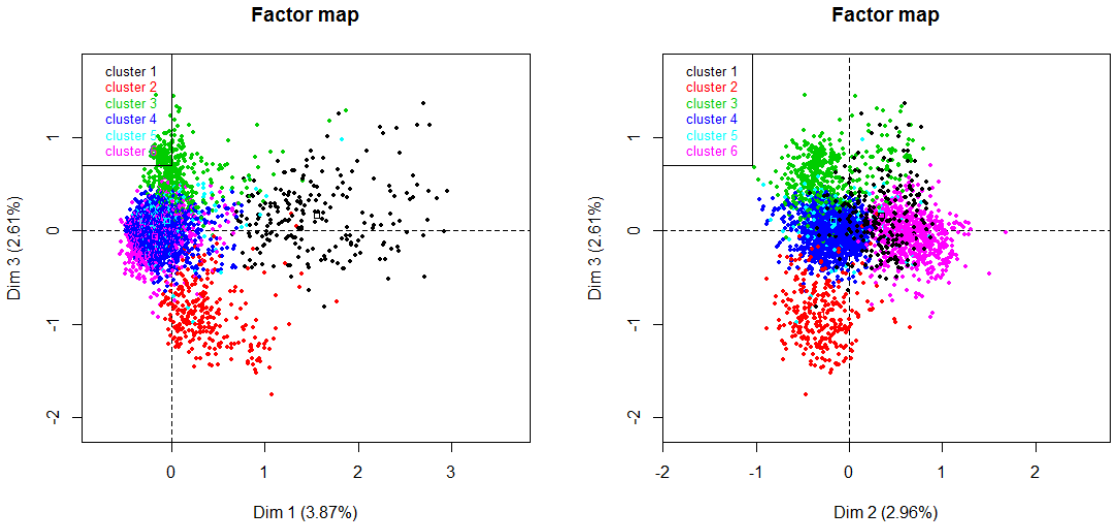


FIGURE 7.18 – 6 classes dans divers dimensions de la MCA.

Ceci clôture la partie sur la classification par MCA de toutes les variables sans DM. Nous allons maintenant développer la méthode intuitive avant d'interpréter et de comparer les deux méthodes.

7.2.2 Intuitivement

Après avoir créé la matrice intuitive associée au même échantillon pris sur l'ensemble des variables sans aucune donnée manquante, nous avons à nouveau procédé, avec R, à une classification ascendante hiérarchique. Les différents dendrogrammes obtenus n'étaient à nouveau pas clairs pour le lien simple, le lien complet, le centroïde et la médiane. L'indice d'agrégation retenu sera celui de Ward, dont le dendrogramme se trouve à la FIGURE 7.19. Ce dendrogramme se coupe facilement en 2 ou 3 classes. Les paliers en-dessous de ce nombre de classes ne sont pas facilement identifiables, puisque plusieurs choses se groupent à des hauteurs différentes.

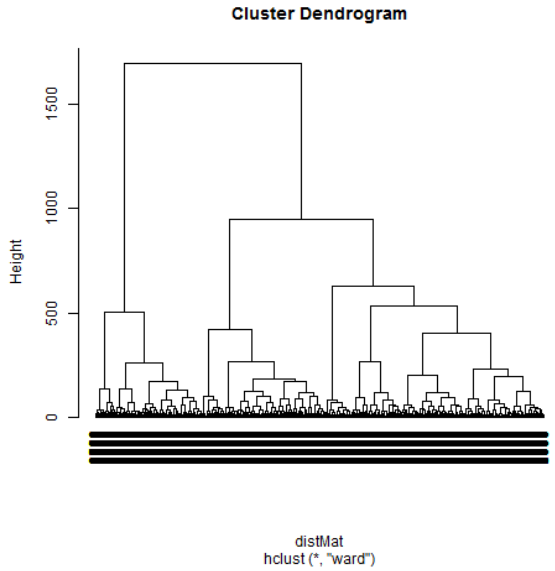


FIGURE 7.19 – Dendrogramme correspondant à la distance intuitive avec toutes les variables sans les DM

Nous allons, par conséquent, retenir les partitions en deux et trois classes, dont la projection dans les deux premières dimensions de la MCA se trouve à la FIGURE 7.20. Nous pouvons voir que la partition en deux classes ne sépare pas les petites valeurs de la dimension 1 des autres, comme dans la classification avec la MCA, mais celles de la dimension 2. Rappelons que cette dimension est principalement guidée par le motif de la consultation.

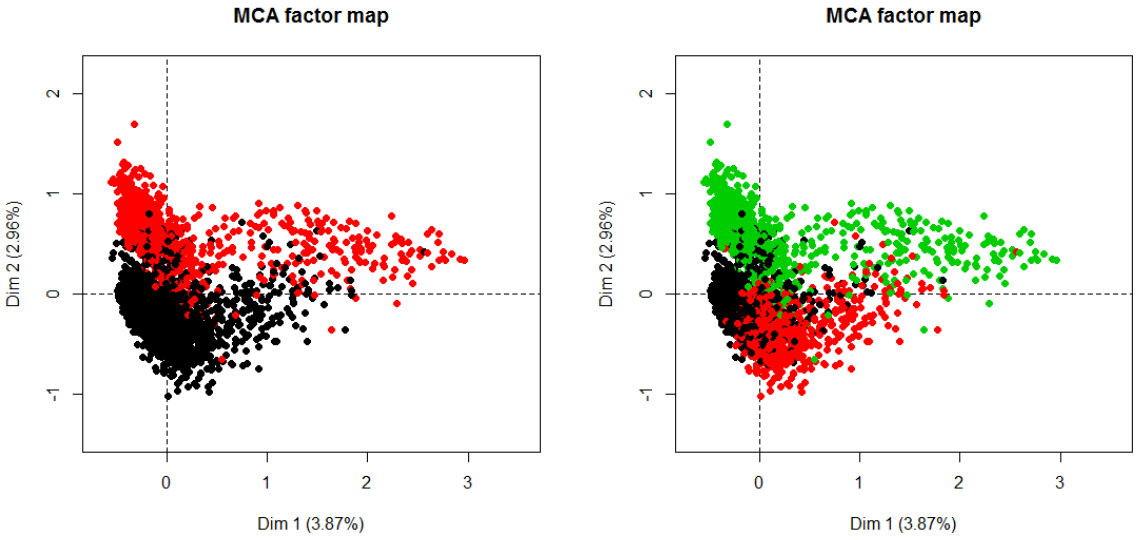


FIGURE 7.20 – Partition intuitive en 2 et 3 classes avec toutes les variables sans les DM

Même si l'algorithme de classification est effectué sur une matrice de distance qui compte simplement le nombre de variables, pour lesquelles les patients n'ont pas la même réponse, ces classes se distinguent aisément dans la MCA. La partition en trois classes divise la classe noire en deux.

7.2.3 Interprétations

Cette fois, les deux méthodes semblent séparer les classes d'une manière différente. À nouveau, les deux méthodes indiquent une classification en deux classes. La méthode intuitive nous donne une classe avec 2956 éléments et l'autre avec 919, alors que la méthode par la MCA a une partie de 3617 patients et l'autre de 258. Dans les deux cas, une classe est beaucoup plus grande que l'autre. Pour pouvoir comparer les classifications, nous allons regarder la répartition des variables au sein de chaque classe. Ainsi, en sommant les proportions de modalités d'une même classe, nous retrouvons 100%. Remarquons que les résultats bruts consistaient à nommer "classe1" celle avec la majorité de données pour l'intuitive, alors que la plus grosse après MCA était appelée "classe2". Nous avons décidé d'appeler, dans les deux cas, "classe1" la plus grosse.

Nous allons effectuer les histogrammes des variables pour lesquelles les tendances dans les classes sont différentes. Après avoir observé tous les graphiques, nous avons conclu que la nationalité, le sexe, la catégorie professionnelle, la province, la classe d'âge, les ressources principales et la nature de la démarche n'étaient pas des données significatives dans la création des classes.

Les deux seules variables socio-économiques qui semblent légèrement différentes dans les deux classes sont la source principale de revenu et le mode de vie. Nous allons voir que, malgré tout, ces différences ne sont pas énormes. Commençons par le revenu à

la FIGURE 7.21. Nous constatons que lorsque nous sommes passés par la MCA, nous n'avons pas de grandes différences entre les deux classes. La méthode intuitive marque une légère tendance pour chaque classe. La classe 1 concerne proportionnellement plus de personnes touchant une pension, sans revenu ou recevant des allocations de maladie et d'invalidité, alors que la classe 2 comprend plutôt les patients avec un revenu propre ou le chômage.

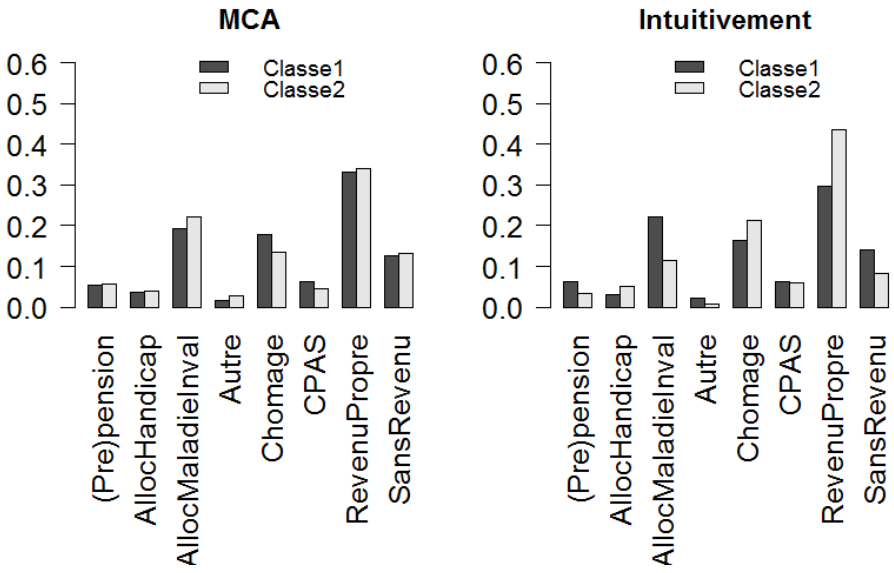


FIGURE 7.21 – Source principale de revenu

À la FIGURE 7.22, au niveau du mode de vie, les deux classes n'ont pas non plus de comportement très divergent.

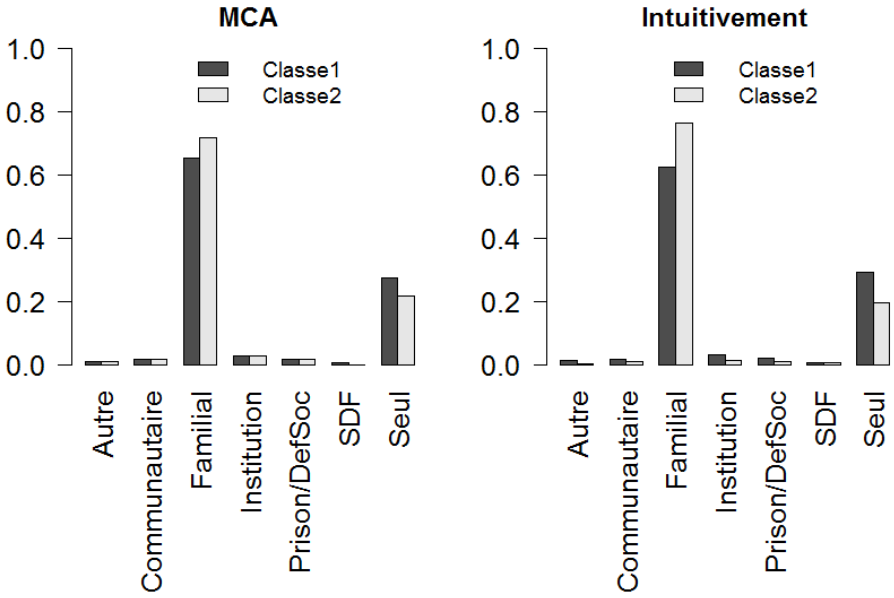


FIGURE 7.22 – Mode de vie

Le classe 2 reprend un peu plus de patients issus d'un milieu familial et la 1 de personnes vivant seules. Dans l'ensemble, les données socio-économiques pures n'influencent pas beaucoup les classes.

Nous pouvons maintenant passer aux variables concernant la consultation, du point de vue du patient. Tout d'abord, nous avons le type de dossier à la FIGURE 7.23. La majorité des personnes ont des dossiers individuels. Malgré tout, dans la partition intuitive, nous avons dans la classe 2 presque tous les dossiers de couple et de famille, complétées par des dossiers individuels. Avec la MCA, cette tendance est moins prononcée.

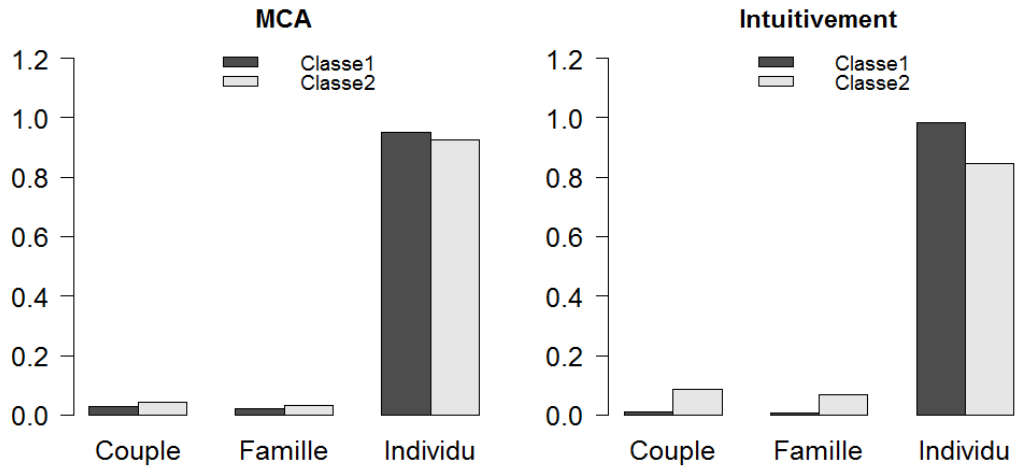


FIGURE 7.23 – Type de dossier

Pour les demandes des patients, nous avons inséré les histogrammes à la FIGURE 7.24. Puisque dans l'ensemble des données, nous avons une grande majorité de personnes qui veulent un suivi, les autres modalités sont également sous-représentées dans chaque classe. Cependant, selon la méthode intuitive, la classe 1 possède proportionnellement plus de personnes voulant un suivi ou une prescription médicale, et la classe 1 comprend les autres demandes. Pour la méthode MCA, aucune tendance n'est à mentionner pour les demandes des patients.

La dernière donnée concernant le point de vue du patient est le motif qui l'a amené à consulter. Nous avions vu précédemment que la méthode par la MCA séparait selon la première dimension, alors que l'intuitive le faisait selon la deuxième. De plus, en observant les carrés des cosinus, nous avions constaté que le motif guidait principalement cette deuxième dimension. Nous avons les histogrammes pour les niveaux des motifs, ainsi que les motifs spécifiques à la FIGURE 7.25 et ils confirment que la classification intuitive est guidée par le motif, alors que l'autre, par la MCA, ne l'est pas. En effet, la classe 1 intuitive comprend plus de 90% des niveaux de motifs "problèmes personnels" et la classe 2 plus de 80% des "problèmes relationnels". Les autres motifs et les consultations sans motif se retrouvent également dans la deuxième classe. Si nous regardons les motifs spécifiques, nous ne pouvons dégager aucune tendance de la méthode des MCA, alors que l'intuitive insère une barrière nette entre les différents types de problèmes.

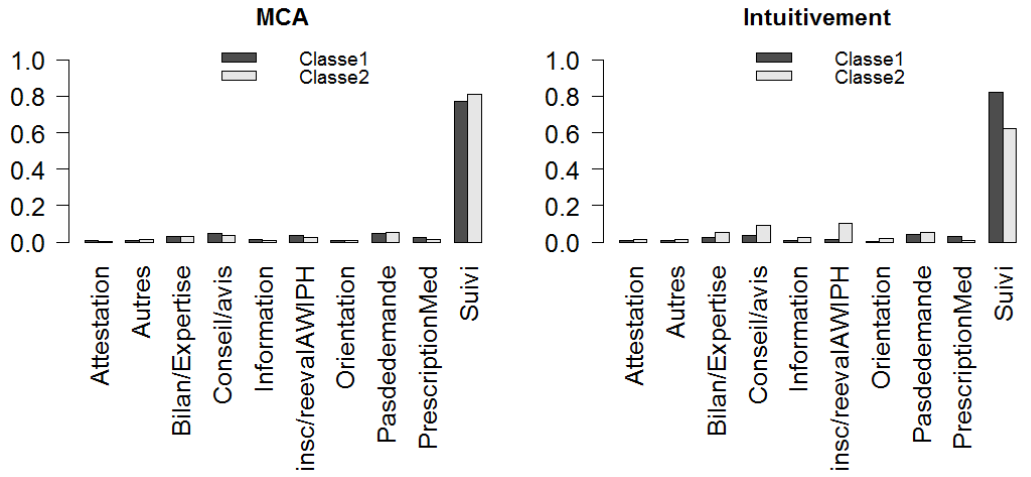


FIGURE 7.24 – Demande du patient

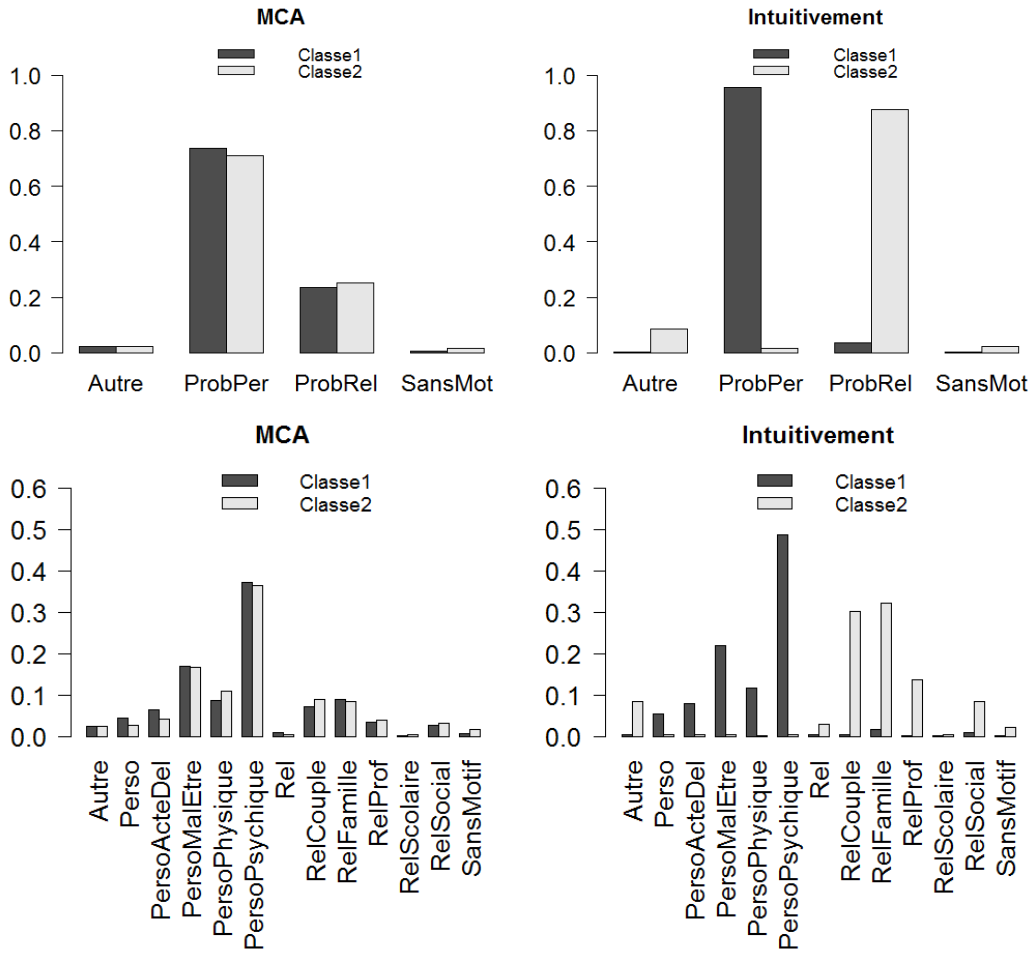


FIGURE 7.25 – Motif principal de la consultation

Pour terminer cette section, passons aux propositions de prise en charge et aux codes ICD10 principaux. Les propositions de prise en charge principale se trouve à la FIGURE 7.26. La méthode des MCA ne donne à nouveau que très peu de différences, alors que la matrice intuitive nous mène à des résultats avec légèrement plus de dissemblances. Nous noterons que la classe 1 a plus de personnes qui se sont vues proposer une thérapie ou un traitement médical.

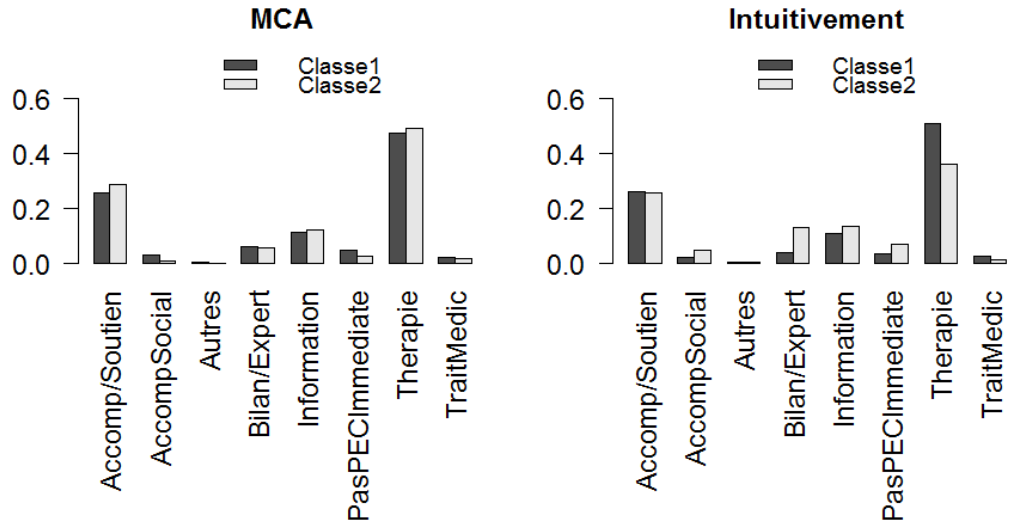


FIGURE 7.26 – Proposition de prise en charge

Enfin, nous avons les codes ICD10 placés en première position à la FIGURE 7.27. Une fois de plus, la méthode par la MCA n'exprime que de légères tendances, alors que la matrice de distance intuitive est plus discriminante. Le plus flagrant est que la classe 2 a une plus grande proportion de codes de type Z. De plus, cette classe possède également plus de type F7 et Autre. Les autres codes sont plus présents dans la classe 1.

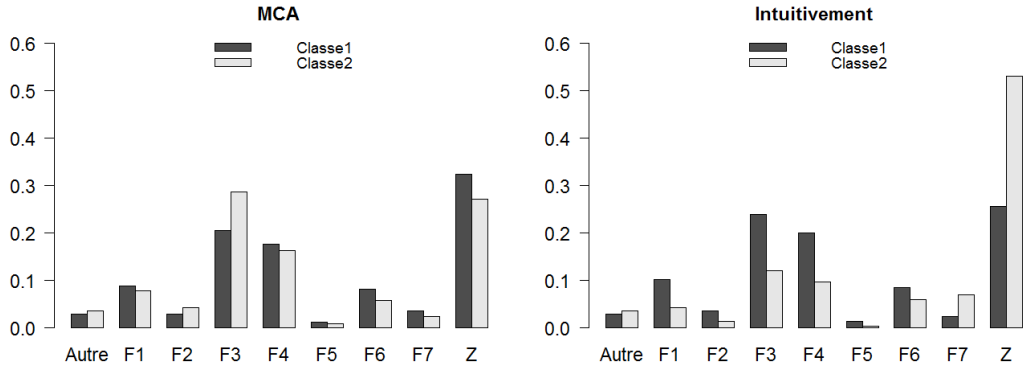


FIGURE 7.27 – Code ICD10 principal

7.2.4 Conclusion

La MCA nous ressort des classes qui ne s'expliquent pas vraiment lorsque nous repassons aux données initiales. Cela est probablement dû à la perte d'information possible, causée par le manque de cohérence entre le critère de la MCA et celui de la classification.

Par contre, au niveau de la méthode plus intuitive, les résultats parlent réellement en terme des variables initiales. Notons qu'il est normal que le motif ait autant d'importance, puisque nous avons conservé à la fois la colonne concernant le niveau du motif et celle du motif spécifique. Par conséquent, quelqu'un qui a un motif personnel a une distance de 2, pour cette donnée, avec un patient de motif relationnel. Cela est à mentionner, mais ne pose aucun problème. En effet, nous voulions qu'entrent en compte le niveau du motif et le motif spécifique, donc il est normal d'avoir gardé les deux colonnes.

La classification obtenue ici est clairement définie par le motif de la consultation. En effet, la classe 1 reprend les problèmes personnels et la classe 2, les problèmes relationnels et autres. La deuxième classe, reprenant les problèmes relationnels et autres, a un plus grand nombre de codes de type Z et F7, donc les facteurs influant sur l'état de santé et les retards mentaux. De plus, les patients de cette classe se sont plus vu proposer principalement des thérapies et traitements médicaux. La première classe, des personnes avec des problèmes personnels, est plus touchée par les autres codes. Les patients de ce groupe ont eu comme prise en charge proposée une information, un accompagnement social, un bilan ou une expertise ou alors n'ont eu aucune proposition de prise en charge immédiate.

Les autres variables n'avaient qu'une légère différence en fonction de la classe. Nous pouvons noter que la classe 1 concerne plus de dossiers individuels, de personnes de mode de vie seul, sans revenu ou avec des allocations maladie et invalidité. La classe 2 comprend les dossiers de couple et de famille, les modes de vie familiaux, et les personnes qui ont comme revenu principal leur revenu propre ou le chômage.

7.3 Conclusion de la classification

En conclusion, nous avons une classification des données socio-économiques si nous ne considérons que ce type de variables. Cette partition dépend de plusieurs données, mais toutes n'ont pas le même poids.

Par contre, si l'on ajoute les codes, les propositions de prise en charge et les motifs spécifiques de la consultation, nous obtenons une classification qui n'est presque plus influencée par les données socio-économiques, mais plutôt par les motifs et les codes.

Au final, on obtient un groupe avec les personnes ayant des problèmes personnels et dont le code principal est F1, F2, F3, F4, F5 ou F6. L'autre classe concerne les personnes avec des problèmes relationnels ou autres et dont les codes sont de type Z ou F7.

Conclusion du mémoire

Ce mémoire aborde deux domaines fort distincts : les statistiques et la santé publique. Nous avons donc dû, dans un premier temps, nous familiariser avec un milieu qui n'était pas du tout celui dans lequel nous travaillions quotidiennement. Le nettoyage des données a été une tâche importante et assez fastidieuse, mais indispensable.

L'ensemble du travail a été effectué avec le soutien du groupe de travail à l'OWS, ainsi que de M. Rémon. Nous avons défini les objectifs ensemble de sorte que les analyses aient un intérêt réel et concret pour l'OWS et soient réalisables au niveau mathématique dans le cadre de ce mémoire. Nous avons d'abord procédé à une analyse descriptive de la population. Ensuite, nous avons recherché des dépendances à l'aide d'échantillonnages aléatoires, car les données étaient trop lourdes pour permettre aux tests d'hypothèses d'être cohérents. Pour terminer, nous avons procédé à une classification des patients.

Les conclusions de ce mémoire sont que les femmes vont plus vite consulter que les hommes. Les hommes ont plus tendance à avoir des démarches contraintes, donc le fait qu'ils soient moins présents ne signifie pas nécessairement que les femmes ont d'avantage de problèmes, mais simplement qu'elles sont plus promptes à consulter. De plus, au niveau des âges, 20 et 40 ans sont des moments où nous sommes plus susceptibles de nous rendre pour la première fois dans un service de santé mentale. Cela peut s'expliquer par le fait que ces deux périodes sont des moments charnières de la vie. Ensuite, la plupart des personnes ont une démarche spontanée et n'ont pas été contraintes de se rendre à la consultation. L'organisme qui a suggéré aux patients de se rendre dans un SSM sont principalement les soins ambulatoires et l'entourage. Nous pouvons ajouter que les ressources non-professionnelles susceptibles d'aider le patient sont essentiellement personnelles et familiales. De plus, les patients ont majoritairement pour diagnostic principal un trouble mental ou un trouble du comportement et en second un diagnostic social. Les traitements médicaux ont été suggérés seulement à 10.2% des patients.

Au niveau du retour de ce mémoire à l'OWS, plusieurs remarques afin d'améliorer le questionnaire ont été suggérées. Par exemple, au niveau du parcours scolaire, les différentes modalités ne permettaient pas de savoir clairement où en était le patient. De plus, pour la langue, il serait plus pertinent de savoir si le patient sait parler français et a donc su parler directement au praticien, sans traducteur, plutôt que seulement sa langue maternelle. Enfin, chaque service avait répondu un peu à sa manière et il a été laborieux d'uniformiser la syntaxe. Cependant, nous savons qu'un travail a déjà été effectué pour améliorer tout cela pour les données de 2012.

Ce travail nous amène a de nouvelles perspectives. Nous pourrions encore faire des cartes par commune et mesurer l'attractivité de chaque service. De plus, au niveau de la classification, une idée pour lutter contre le problème des données qualitatives aurait été de classer d'abord les variables, pour faire alors une MCA au sein de chaque classe et avoir différents prototypes [27].

Notons que ce sujet m'a permis d'avoir une première approche du monde du travail. En effet, ce mémoire consistait à suivre les directives et les exigences du groupe de travail de l'OWS. J'ai pu me rendre compte du mode de fonctionnement des réunions d'équipe où chacun est expert dans son propre domaine. Il a fallu que je m'adapte au langage de la santé publique et que je trouve les mots pour leur expliquer le côté statistique, ainsi que les limites de certaines illustrations. Dans le cadre de cette étude, j'ai également eu la chance d'être invitée à un "midi social-santé", qui consiste en une conférence sur un sujet de santé publique, donnée à l'OWS sur le temps de midi, une fois par mois. Ensuite, en avril, j'ai pu présenter mes résultats lors d'un de ces événements. L'expérience acquise grâce à ce travail est très enrichissante pour moi au niveau professionnel.

Bibliographie

[1] http ://www.soft-concept.com/surveymag/lexique/lexique-etudes-marketing-table-khi2.htm. consultée le 21 avril 2014.

[2] http://www.soins-infirmiers.com/demarche_de_sante_publique.php. consultée le 10 avril 2013.

[3] http://www.douglas.qc.ca/page/reforme-sante-mentale. consultée le 12 avril 2013.

[4] A. Bouchier. La classification ascendante hiérarchique (c.a.h.). 2010. Disponible sur http://rstat.ouvaton.org/?article10/classification-hierarchique.

[5] L'Association canadienne pour la santé mentale Chaudière-Appalaches. http://www.acsm-ca.qc.ca/definition-sm/. consultée le 10 avril 2013.

[6] centre des médias : La santé mentale : renforcer notre action. Aide-mémoire n°220. http://www.who.int/mediacentre/factsheets/fs220/fr/, Septembre 2010. consultée le 10 avril 2013.

[7] OMS (Organisation Mondiale de la Santé). http://www.who.int/topics/mental_health/fr/index.html. consultée le 10 avril 2013.

[8] Agence de la santé et des services sociaux de Montréal (Québec). Santé des populations et services de santé. http://www.dsp.santemontreal.qc.ca/dossiers_thematiques/services_preventifs/thematique/sante_des_populations_et_services_de_sante/. consultée le 12 avril 2013.

[9] Institut Wallon de la Santé Mentale (IWSM). Aux portes du soin : l'accessibilité en santé mentale. http://www.iwsm.be/cahier/Cahier_3.pdf. consultée le 12 avril 2013.

[10] Institut Wallon de la Santé Mentale (IWSM). Service de santé mentale de wallonie (mise à jour : janvier 2010). http://www.iwsm.be/pdf_dir/listessm. consultée le 9 mai 2013.

[11] Institut Wallon de la Santé Mentale (IWSM). Vue d'ensemble. http://www.iwsm.be/institut-wallon-sante-mentale-Vue-Ensemble.php. consultée le 12 avril 2013.

[12] Institut Wallon de la Santé Mentale (IWSM). Fiche d ?enregistrement de données à caractère épidémiologique : Manuel d ?utilisation, 2005. Version n°2.

[13] Institut Wallon de la Santé Mentale (IWSM). Médecine générale et santé mentale. *Confluences (3 revues par an)*, (12), Décembre 2005. Disponible sur http://www.iwsm.be/confluences/C12.pdf.

[14] Institut Wallon de la Santé Mentale (IWSM). Zoom sur les services de santé mentale. *Confluences (3 revues par an)*, (14), Septembre 2006. Disponible sur `http://www.iwsm.be/confluences/C14.pdf`.

[15] Observatoire Wallon de la Santé (OWS). *rapport d'activités*. 2010.

[16] Ministère des affaires sociales et de la Santé (Français). Psychiatrie et santé mentale : 2005-2008. `http://www.sante.gouv.fr/IMG/pdf/plan_2005-2008.pdf`. consultée le 12 avril 2013.

[17] Jeanine Pommier et Olivier Grimaud. Les fonctions essentielles de santÉ publique : Histoire, dÉfinition et applications possibles. *Santé publique*, 19(1) :pp. S9–S14, Janvier-Février 2007. Disponible sur `http://www.cairn.info/revue-sante-publique-2007-hs-page-9.htm`.

[18] Paul S. Levy et Stanley Lemeshow. *Sampling of Populations : Methods and Applications - Third Edition*. 1999.

[19] Julie Hubert et Sylvie Lopez Robillard. `http://www.infirmiers.com/etudiants-en-ifsi/cours/cours-sante-publique-notions-de-base.html`, 2007. consultée le 10 avril 2013.

[20] Isabelle Jeuge-Maynart. *Le petit Larousse illustré*. Larousse, 2007.

[21] Léonce MPUNIKIYE. *Les déterminants et les contextes de la santé et du bien-être publique*. PhD thesis, Faculté des sciences - Unité de statistique (FUNDP), Févier 2012.

[22] Coordination Prof C. Gosset Prof. P. Gillet, Prof. D. Porignon. *Introduction générale à la santé publique*. PhD thesis, Université de Liège : Sciences de la Santé publique, 2009-2010.

[23] Service public de Wallonie (SPW). *Tableau de bord de la santé en Wallonie*. 2009.

[24] LE CAHIER D ?INTERVENTION SPIRITUELLE EN SANTÉ MENTALE (CISSM) (Québec). Les diffÉrents milieux d ?insertion ou points de services. *Révision*, mars 2012. Disponible sur `http://www.iwsm.be/confluences/C14.pdf`.

[25] M. Rémon. Analyse mutlivariée et logiciels statistiques, 2011-2012.

[26] Dr Begue simon A. `http://facmed.univ-rennes1.fr/wkf/stock/RENNES20090116095023ambegueIntroduction_A_la_SantE_Publique.doc`, 2011. consultée le 10 avril 2013.

[27] J. Candau P. Deuffic M. Chavent J. Saracco V. Kuentz-Simonet, S. Lyser. Classification de variables qualitatives pour la compréhension de la prise en compte de l'environnement par les agriculteurs. 2012. Disponible sur Google Scholar.

[28] Wallonie. Guide vers de meilleurs soins en santé mentale. Disponible sur `http://www.psy107.be/SiteFiles/Wallonie.pdf%20def.pdf`.

Table des figures

Annexes

Questionnaire

Mise en page M.R.W. - D.G.A.S.S. 01-2004 I.W.S.M. - Version VIII-01-2004

Fiche d'enregistrement de données à caractère épidémiologique

Fiche "Adultes" (18 ans et +)
(Une fiche par dossier et pour tout nouveau dossier)

3. Année d'enregistrement ☐☐☐☐ 6 Type de dossier ☐

7. Né(e) en ☐☐☐☐ 8. Sexe ☐ 9. Domicile (code postal) ☐☐☐☐

17. Nature de la démarche ☐ 18. Origine de la démarche ☐☐☐☐ + ☐☐☐☐

10. Etat civil ☐
1 célibataire 5 divorcé
2 marié 6 veuf
3 séparé 8 inconnu
4 contrat de vie commune

11. Nationalité ☐
1 belge 2 autre UE
3 hors UE 8 inconnu

12. Langue maternelle ☐
1 français 5 italien
2 néerlandais 6 arabe
3 allemand 7 turc
4 anglais 8 inconnu
9 autre : .....

13. Mode de vie ☐☐
10 seul
20 familial
21 en couple
22 partenaire + enfant(s)
23 avec les enfants
24 chez les enfants
25 chez les parents
26 famille recomposée
27 avec 1 autre membre de la famille
29 autre: ....
30 communautaire
40 institution d'aide ou de soins
50 sans domicile fixe
60 prison/défense sociale
80 inconnu
90 autre:
.....

14. Niveau de scolarité

A. Atteint ☐☐ B. Dernière année réussie ☐☐

	Selon l'année réussie
10 pas suivi d'enseignement	
20 maternel	=> 21 22 23
30 primaire	=> 31 32 33 34 35 36
31 normal	
32 spécial	
40 secondaire	=> 41 42 43 44 45 46
41 général	
42 technique	
43 professionnel	49 certificat de qual.
44 spécial	de l'ens. Spécial
50 supérieur non universitaire	=> 51 52 53 54
60 universitaire	=> 61 62 63 64 65 66+
70 promotion sociale	
75 contrat d'apprentissage	=> 71 72 73 74
80 inconnu	90 autre:
90 autre:	

.....
.....

15. Catégorie professionnelle ☐
1 ouvrier 6 étudiant
2 employé 7 sans profession
3 cadre/directeur 8 inconnu
4 profession libérale 9 autre:
5 indépendant .....

16. Source principale de revenus ☐
1 propre activité prof 7 sans revenus
2 allocations de chômage 8 inconnu
3 (pré-)pension 9 autre:
4 allocation de handicap .....
5 allocation maladie/invalidité
6 allocations du CPAS

19. Prises en charge antérieures: ☐☐☐☐ + ☐☐☐☐ + ☐☐☐☐ (1)

(1) sélection. les codes parmi ceux du tableau général des types professionnels et des services (cf. annexe du manuel d'utilisation ou la fiche jaune).

Questionnaire

20. Demande du consultant ☐☐

10 suivi
- 11 thérapie
- 12 rééducation
- 13 soutien
- 14 accompagnement social
- 19 autre: .....

20 bilan/expertise
30 inscription/réévaluation AWIPH
40 autres demandes
- 41 information
- 42 conseil/avis
- 43 orientation
- 44 attestation
- 45 prescription médicale

50 pas de demandes précises
80 inconnu
90 autres : .....

21. Motifs présentés lors de la première consultation ☐☐ +(☐☐)

10 problématiques personnelles
- 11 plaintes et symptômes physiques
- 12 plaintes et symptômes psychiques
- 13 mal-être
- 14 acte(s) délictueux
- 19 autres: .....

20 problématique relationnelle
- 21 difficultés ppales dans le couple
- 22 difficultés ppales dans le milieu familial
- 23 difficultés ppales dans le milieu social
- 24 difficultés ppales dans le milieu professionnel
- 25 difficultés ppales dans le milieu scolaire
- 29 autre: .....

30 sans motif
80 inconnu
90 autre: .....

22. Codes ICD10 :

☐☐☐ · ☐☐ intitulé: .....
☐☐☐ · ☐☐ intitulé: .....
☐☐☐ · ☐☐ intitulé: .....
☐☐☐ · ☐☐ intitulé: .....
☐☐☐ · ☐☐ intitulé: .....
☐☐☐ · ☐☐ intitulé: .....

23. Proposition de prise en charge ☐☐ + ☐☐ + ☐☐

10 information/clarification
20 thérapie
- 21 individuelle
- 22 familiale
- 23 de couple
- 24 de groupe

30 accompagnement et soutien
- 31 individuel
- 32 familial
- 33 de couple
- 34 de groupe

40 accompagnement social
- 41 individuel
- 42 familial
- 43 de couple
- 44 de groupe

50 rééducation
60 bilan/expertise
70 traitement médicamenteux
80 pas de prise en charge immédiate
- 81 orientation vers : =======> ☐☐☐☐ (1)
- 82 liste d'attente
- 83 aucune proposition
- 89 autre: .....

90 autre: .....

24. Réseau professionnel (1)

☐☐☐☐ ☐☐☐☐
☐☐☐☐ ☐☐☐☐

25. Ressources ☐☐ + ☐☐ + ☐☐

10 personnelles
20 familiales
30 amis
40 entourage social
50 entourage professionnel
80 inconnu
90 autre: .....

26. Variables à usage interne

A ☐ B ☐ C ☐ D ☐ E ☐☐ F ☐☐

G ☐☐ H ☐☐ I ☐☐☐☐ J ☐☐☐☐

K ☐☐☐☐ L ☐☐☐☐

(1) voir note page précédente.

Etablissements présents dans les données

Etablissement	Etablissement	Etablissement
AICS(Dinant)	Flemalle	MorlanwelzAriane
AIGS	Florennes	Mouscron
Andenne	Gembloux	NamurAstrid(Exil)
Angleur	Gosselies	NamurAstridANA
Arlon	Huy	NamurAstridGeneraliste
AthHoton	Jambes	NamurBalances
AthLaPasserelle	Jambesmissionsurdite	Nivellesbxl
Bastogne	Jemelle	Nivellescpas
Beauraing	Jemeppe	Ottignies
Binche	Jodoigne	Ougree
Bouillon	Jolimont	Seraing
BoussuLaKalaude	LaLouvière	Soignies
Brainel'Alleud	LeRoeulxLeDièse	StChristophe
CharleroiBernus	Libramont	StGhislain
CharleroiCPASAICS	LibramontAICS	StHubert
CharleroiCPASmissionGeneraliste	LiègeAlfa	Tamines
CharleroiCPASTox	LiègeBaillon	Tournaiathenee
CharleroiCPASTrialogue	LiègeBaillonMSTabane	TournaiBeyaert
CharleroiGrandRue	LiègeBouvy	Tubize
CharleroiScience	LiègeFranchimontois	Verviersdeportes
Châtelet	LiègePsychoJ	VerviersDinant
Chimay	LiègeReversSiajef	VerviersDinantCentreII
Ciney	LiègeUnif	VerviersDinantSAPI
ClubTheoVanGogh	LLN	Virton
Colfontaine	Lobbes	ViseSluse
Comines	Malmedy	Wavre
Courcelles	Marche	
Couvin	MarchienneauPont	
Dinant	Mons	

Table 7.3 – Les services de santé mentale présents dans la base de données

Liste des services de santé mentale agréés de l'Institut Wallon de la Santé Mentale [10]

Service de Santé Mentale de Wallonie (mise à jour : janvier 2010)

N°	NOM	ADRESSE	CP	LOCALITE	PROVINCE	TELEPHONE
Province de NAMUR						
001/1	SERVICE DE SANTE MENTALE D'ANDENNE	Rue de l'Hôpital, 23	5300	ANDENNE	NAMUR	085/84.94.90
016/2	SERVICE DE SANTE MENTALE DE BEAURAING	Rue de l'Aubépine, 61	5570	BEAURAING	NAMUR	082/71.47.51
015/1	SERVICE DE SANTE MENTALE DE COUVIN	Ruelle Cracsot, 12	5660	COUVIN	NAMUR	060/34.52.33
016/1	SERVICE DE SANTE MENTALE DE DINANT	Rue Daoust, 72	5500	DINANT	NAMUR	082/21.48.20
015/2	SERVICE DE SANTE MENTALE DE FLORENNES	Rue Gérard de Cambrai, 18	5620	FLORENNES	NAMUR	071/68.10.21
045/2	SERVICE DE SANTE MENTALE DE GEMBLOUX	Rue Albert 1er, 3	5300	GEMBLOUX	NAMUR	081/62.66.26
021/1 et 021/2	SERVICE DE CONSULTATIONS PSYCHO-SOCIALES pour ENFANTS et ADULTES	Rue de Dave, 124	5100	JAMBES	NAMUR	081/30.55.20
	SERVICE DE SANTE MENTALE POUR PERSONNE SOURDES	Rue de Dave, 124	5100	JAMBES	NAMUR	0498/26.08.62
016/3	SERVICE DE SANTE MENTALE DE JEMELLE	Rue de Ninove, 32	5580	JEMELLE	NAMUR	084/34.42.26
	SERVICE DE SANTE MENTALE EXIL	Rue Docteur Haibe, 4	5000	NAMUR	NAMUR	081/73.67.22
037/1	CENTRE PSYCHOTHERAPIQUE PROVINCIAL	Avenue Reine Astrid, 20A	5000	NAMUR	NAMUR	081/77.67.13
038/1	SERVICE DE SANTE MENTALE PROVINCIAL	Rue Château des Balances, 3B	5000	NAMUR	NAMUR	081/77.67.12
	SSM "AVEC NOS AINES"	Rue Martine Bourtonbourt, 2	5000	NAMUR	NAMUR	081/72.95.82
	SERVICE DE SANTE MENTALE	Rue Duculot, 11	5060	TAMINES	NAMUR	071/26.99.10
Province du BRABANT WALLON						
007/1	SERVICE DE SANTE MENTALE DE BRAINE L'ALLEUD "LE SAFRAN"	Rue Jules Hans, 43	1420	BRAINE L'ALLEUD	BRABANT WALLON	02/384.68.46
040/2	SERVICE DE SANTE MENTALE DE JODOIGNE	Chaussée de Tirlemont, 89	1370	JODOIGNE	BRABANT WALLON	010/81.31.01
055/1	SERVICE DE SANTE MENTALE	Grand Place, 43	1348	LOUVAIN LA NEUVE	BRABANT WALLON	010/47.44.08
055/2	SERVICE DE SANTE MENTALE	Rue Paulin Ladeuze, 7	1348	LOUVAIN LA NEUVE	BRABANT WALLON	010/47.40.14
040/1	CG DE LA PROVINCE DU BRABANT	Chaussée de Bruxelles, 55	1400	NIVELLES	BRABANT WALLON	067/21.91.24
040/2	SERVICE DE SANTE MENTALE DU CPAS	Rue Samiette, 70	1400	NIVELLES	BRABANT WALLON	067/28.11.50
042/1	CENTRE MEDICO-PSYCHOTHERAPIQUE	Rue des Fusillés, 20	1340	OTTIGNIES	BRABANT WALLON	010/41.88.78
040/3	SERVICE DE SANTE MENTALE "LA FORGE DE VIE"	Rue de Château, 42	1480	TUBIZE	BRABANT WALLON	02/390.06.37
054/2	SERVICE DE SANTE MENTALE DU BRABANT-WALLON	Avenue du Belloy, 45	1300	WAVRE	BRABANT WALLON	010/22.54.03
Province du HAINAUT						
003/1	SERVICE DE SANTE MENTALE "LA PASSERELLE"	Square Saint Julien, 43	7800	ATH	HAINAUT	068/28.55.01
004/1	SERVICE DE SANTE MENTALE PSYCHOLOGIQUE de ATH	Rue Isidore Hoton, 9	7800	ATH	HAINAUT	068/26.50.90
006/1	SERVICE DE SANTE MENTALE PSYCHOLOGIQUE DE BINCHE	Rue de Bruxelles, 18	7130	BINCHE	HAINAUT	064/33.63.68
010/1	ACCUEIL MEDICO-PSYCHOLOGIQUE	Grand Rue, 67	6000	CHARLEROI	HAINAUT	071/70.00.03
011/1	CENTRE DE SANTE MENTALE DU CPAS	Rue Léon Bernus, 18	6000	CHARLEROI	HAINAUT	071/32.94.18
011/2	CLUB PSYCHO-SOCIAL THEO VAN GOGH	Rue Destrée, 45	6000	CHARLEROI	HAINAUT	071/28.19.47
009/1	SERVICE DE SANTE MENTALE DE CHARLEROI	Rue Léon Bernus, 22	6000	CHARLEROI	HAINAUT	071/31.63.78
008/1	SERVICE DE SANTE MENTALE DE CHARLEROI	Rue de la Science, 7	6000	CHARLEROI	HAINAUT	071/20.72.80
	SSM DU CPAS "ESPACE SANTE"	Boulevard Zoé Drion, 1 (7ème Et.)	6000	CHARLEROI	HAINAUT	
012/1	CENTRE D'ACCUEIL PSYCHO-SOCIAL	Rue du Collège, 39	6200	CHATELET	HAINAUT	071/38.46.38
	SERVICE DE SANTE MENTALE DE CHIMAY "LE PORTAIL"	Chaussée de Trélon, 1	6460	CHIMAY	HAINAUT	060/41.23.54
013/1	SERVICE DE SANTE MENTALE DE COLFONTAINES	Rue Maubeuge, 7	7340	COLFONTAINE (Wasmes)	HAINAUT	065/71.10.30
036/2	SERVICE DE SANTE MENTALE	Chaussée de Warneton, 20	7780	COMINES	HAINAUT	056/55.71.51
014/1	SERVICE DE SANTE MENTALE DE COURCELLES	Rue de la Croisette, 109	6180	COURCELLES	HAINAUT	071/46.60.80
018/1	SERVICE DE SANTE MENTALE DE JOLIMONT	Rue Ferrer, 196	7100	HAINE SAINT PAUL	HAINAUT	064/23.33.53
(023/1)	SERVICE DE SANTE MENTALE DE LA LOUVIERE "PSY CHIC"	Rue du Moulin, 79	7200	LA LOUVIERE	HAINAUT	064/22.25.71
018/2	SERVICE DE SANTE MENTALE "LE PICHOTIN"	rue Albert 1er, 21	6540	LOBBES	HAINAUT	071/55.92.30
033/1	SERVICE DE SANTE MENTALE "LA PIOCHE"	Rue Royale, 95	6030	MARCHIENNE AU PONT	HAINAUT	071/31.18.92
057/1	SERVICE DE SANTE MENTALE "LE DIESE"	Rue des Déportés, 7	7070	MIGNAULT	HAINAUT	067/21.24.77
034/1	SERVICE DE SANTE MENTALE DE MONS	Avenue d'Hyon, 45	7000	MONS	HAINAUT	065/35.43.71
	SERVICE DE SANTE MENTALE "LE PADELIN"	Avenue Baudouin de Constantinople, 2	7000	MONS	HAINAUT	065/357178
036/1	SERVICE DE SANTE MENTALE	Rue de la Station, 161	7700	MOUSCRON	HAINAUT	056/34.67.89
056/1	SERVICE DE SANTE MENTALE	Rue de l'Abbaye, 29/31	7730	SAINT- GHISLAIN	HAINAUT	065/46.54.06
044/1	SERVICE DE SANTE MENTALE ET DE PSYCHOLOGIE	Ruelle Scaffart	7060	SOIGNIES	HAINAUT	067/33.10.68
047/1	SERVICE DE SANTE MENTALE DU TOURNAISIS	Rue Beyaert, 59B	7500	TOURNAI	HAINAUT	069/22.05.13
047/2	SERVICE DE SANTE MENTALE DU TOURNAISIS	Avenue Bozière, 3	7500	TOURNAI	HAINAUT	069/22.30.28
	SERVICE DE SANTE MENTALE (UPPL)	rue Despars, 92	7500	TOURNAI	HAINAUT	069/88.83.33
046/1	SERVICE DE SANTE MENTALE PSYCHOLOGIQUE	Rue de l'Athénée, 21	7500	TOURNAI	HAINAUT	069/22.72.48

Liste des services de santé mentale agréés de l'Institut Wallon de la Santé Mentale [10]

N°	NOM	ADRESSE	CP	LOCALITE	PROVINCE	TELEPHONE
Province de LIEGE						
030/2	SERVICE DE SANTE MENTALE D'ANGLEUR "SOSL"	Rue Vaudée, 40	4031	ANGLEUR	LIEGE	04/365.14.53
039/1	SERVICE DE SANTE MENTALE DE COMBLAIN AU PONT	Route d'Aywaille, 22	4170	COMBLAIN AU PONT	LIEGE	04/369.23.23
	SOZIAL PSYCGOLOGISCHES ZENTRUM (SPZ)	Verviersstrasser, 14 (2ème étage)	4700	EUPEN	LIEGE	087/59.80.59
017/1	SERVICE DE SANTE MENTALE DE FLEMALLE	Rue Spinette, 2	4400	FLEMALLE	LIEGE	04/235.10.50
053/2	SERVICE DE SANTE MENTALE DE HANNUT	Rue J. Mottin, 6	4280	HANNUT	LIEGE	019/51.29.66
019/1	SERVICE DE SANTE MENTALE DE HERSTAL	Rue Lambert, 84	4040	HERSTAL	LIEGE	04/240.04.08
019/2	SERVICE DE SANTE MENTALE DE HERSTAL	Rue Lambert, 86	4040	HERSTAL	LIEGE	04/248.48.10
020/1	SERVICE DE SANTE MENTALE "L'ACCUEIL"	Rue de la Fortune, 6	4500	HUY	LIEGE	085/25.42.26
043/2	SERVICE DE SANTE MENTALE DE JEMEPPE	Voie du Promeneur, 13	4101	JEMEPPE SUR MEUSE	LIEGE	04/231.10.42
022/2	SERVICE DE SANTE MENTALE DE JUPILLE	Cité André Renard, 15	4020	JUPILLE	LIEGE	04/365.12.37
027/1	CENTRE DE REEDUCATION DE L'ENFANCE	Rue En Hors Château, 59	4000	LIEGE	LIEGE	04/223.55.08
029/1	CENTRE DE SANTE MENTALE ENFANTS-PARENTS	Rue Lambert Lebegue,16	4000	LIEGE	LIEGE	04/223.41.12
028/1	CLUB ANDRE BAILLON	Rue des Fontaines Roland, 7-9	4000	LIEGE	LIEGE	04/221.18.50
028/2	CLUB ANDRE BAILLON II	rue Louis Jamme, 29	4000	LIEGE	LIEGE	04/342.97.11
non agréé (030/3	SERVICE DE SANTE MENTALE	Place St-Christophe, 2	4000	LIEGE	LIEGE	04/221.18.50
026/1	SERVICE DE SANTE MENTALE "LE CLIPS"	Rue Alex Bouvy, 18	4000	LIEGE	LIEGE	04/341.29.66
025/1	SERVICE DE SANTE MENTALE ALFA	rue de la Madeleine, 17	4000	LIEGE	LIEGE	04/223.09.03
030/1	SERVICE DE SANTE MENTALE DE LIEGE	Rue des Franchimontois, 4b	4000	LIEGE	LIEGE	04/227.36.17
	SERVICE DE SANTE MENTALE "REVERS-SIAJEF"	rue Maghin, 18-19	4000	LIEGE	LIEGE	04/227.68.76
031/1	SERVICE DE SANTE MENTALE "L'ESPOIR"	Rue Derrière les Murs, 5	4960	MALMEDY	LIEGE	080/33.81.65
065/1	SERVICE DE SANTE MENTALE	Chaussée Churchill, 28	4420	MONTEGNEE	LIEGE	04/364.06.85
039/2	SERVICE DE SANTE MENTALE DE NANDRIN	rue Albert Botty, 1	4550	NANDRIN	LIEGE	085/51.24.15
043/3	SERVICE DE SANTE MENTALE	Rue Berthelot, 29	4102	OUGREE	LIEGE	04/337.49.53
052/1	SERVICE DE SANTE MENTALE	rue du Ponçay, 1	4680	OUPEYE	LIEGE	04/264.33.09
	BERATING UND LEBENSHIEFE - SP2	Wiesenbachstrasse, 5	4780	SAINT-VITH	LIEGE	080/22.76.18
043/1	SERVICE DE SANTE MENTALE DE SERAING-OUGREE	Rue Hya, 71	4100	SERAING	LIEGE	04/337.20.64
022/1	SERVICE DE SANTE MENTALE	Rue de l'Egalité, 250	4630	SOUMAGNE	LIEGE	04/377.46.65
048/1	CENTRE FAMILIAL D'EDUCATION	Rue des Déportés, 30	4800	VERVIERS	LIEGE	087/22.13.92
049/3	SERVICE DE SANTE MENTALE	Rue de Dinant, 22	4800	VERVIERS	LIEGE	087/22.16.84
049/2	SERVICE DE SANTE MENTALE	Rue de la Maison Communale, 4	4802	VERVIERS (HEUSY)	LIEGE	087/22.88.44
049/1	SERVICE DE SANTE MENTALE DE VERVIERS	Rue de Dinant 18-20	4800	VERVIERS	LIEGE	087/22.16.45
051/1	CENTRE FAMILIAL D'EDUCATION "LEON HALKEIN"	Rue de Sluse, 17	4600	VISE	LIEGE	04/379.39.36
050/1	SERVICE DE SANTE MENTALE	Rue de la Fontaine, 53	4600	VISE	LIEGE	04/379.32.62
053/1	SERVICE DE SANTE MENTALE	Rue G. Joachim, 49	4300	WAREMME	LIEGE	019/32.47.92
Province de Luxembourg						
002/2	CLUB DE JOUR d'ARLON	Avenue de Luxembourg, 8	6700	ARLON	LUXEMBOURG	063/22.68.90
002/1	SERVICE DE SANTE MENTALE D'ARLON	Rue Léon Casthillon, 62	6700	ARLON	LUXEMBOURG	063/22.15.54
005/1	SERVICE DE SANTE MENTALE DE BASTOGNE	Rue des Scieries, 71	6600	BASTOGNE	LUXEMBOURG	061/21.28.08
024/2	SERVICE DE SANTE MENTALE DE BOUILLON	Rue du Collège, 5	6830	BOUILLON	LUXEMBOURG	061/46.76.67
024/1	SERVICE DE SANTE MENTALE DE LIBRAMONT	Grand-Rue 8	6800	LIBRAMONT	LUXEMBOURG	061/22.38.72
032/1	SERVICE DE SANTE MENTALE DE MARCHE	Rue du Luxembourg, 15	6900	MARCHE EN FAMENNE	LUXEMBOURG	084/31.20.32
024/3	SERVICE DE SANTE MENTALE	Rue du Mont, 87	6870	ST HUBERT	LUXEMBOURG	061/61.16.20
	SERVICE DE SANTE MENTALE	rue du Docteur Jeanty, 4	6760	VIRTON	LUXEMBOURG	063/45.78.62

Liste des informations récoltées

Nom de la variable	Information concernée
EtablissementXX	Code de l'établissement
AnnéeEnregistrement	Année d'enregistrement du nouveau patient
TypeDossier	Type du dossier qui peut être "en couple", "individuel" ou "Famille"
AnneeNaissance	Année de naissance du patient
Sexe	Sexe du patient
CodePostal	Code postal du patient
EtatCivil	Etat civil du patient
Nationalite	Nationalité du patient
LangueMaternelle	Langue maternelle du patient
ModedevieNiveau	Mode de vie du patient ;"La situation de vie courante" ([12]) Complémentaire de l'état civil.
NiveauScolaritéAtteint	Plus haut niveau d'étude atteint (primaire, secondaire, supérieur)
NiveaudeScolarité-Atteint(Spec)	Type d'étude de ce plus haut niveau (Professionnel, général,..)
NiveaudeScolarité-DernièreAnnéeRéussie	Dernière année réussie au sein de ces études
CatégorieProfessionnelle	Catégorie professionnel du patient
SourcePrincipaledeRevenus	Source principale du revenu (Revenu propre, CPAS, pensions,...)
NatureDémarche	Nature de la démarche du patient (spontanée, orientée ou contrainte)
OriginedelaDémarche	Organisme qui a aidé le patient à se décider
Prise-en-Charges-Antérieures	Réseau professionnel ayant pris le patient en charge préalablement
DemandeduConsultant	Attentes du patient de sa visite (un suivi, une prescription médicale,...)
MotifsNiveau	Niveau du motif de la première consulation (relationnel ou personnel)
MotifsPremièreConsultation	Motif spécifique de la consultation
Code-1a, Code-2a, Code-3a	Classification des pathologies en fonction d'un code nommé "ICD-10"
PropositiondePriseenCharge	Proposition de prise en charge (thérapie, traitement médical, ...)
RéseauProfessionnel	Réseau professionel concerné par la proposition de prise en charge
Ressources	Ressources du patient en terme de soutien extra-professionnel

Liste des types de réseaux professionnels [12]

1000 Entourage 1001 Parent / famille 1002 Ami/voisin/relation 1003 Consultant du centre 1090 Autre :……..	**1400 Aide à la jeunesse** 1401 SAJ 1402 SPJ 1403 COE 1404 AMO 1405 SPEP 1406 IPPJ 1407 Serv. d'accueil résidentiel 1490 Autre:……	**1700 Soins de santé résidentiels** 1701 Hôpital général 1702 Service d'urgence 1790 Autre:…… **1800 Troisième âge** 1801 Maison de repos 1802 Maison de repos et de soins 1890 Autre:……..	**2100 Santé mentale intra-muros** 2101 Serv. d'urgence psy 2102 Serv. psy d'Hôp.G. 2103 Service K jour/nuit 2104 Service K de jour 2105 Hôp. psychiatrique 2106 Hôpital de jour 2107 Habitation protégée 2108 Comm.thérapeutiq. 2109 Struct. Toxico 2110 M.S.P. 2190 Autre:……
1100 Milieu scolaire 1101 Enseignant 1102 PMS 1103 PSE 1104 Centre format°prof. 1105 Logopède ds cadre scol. 1106 Directeur d'école 1107 Ecole de devoirs 1190 Autre :……	**1500 Justice / police** 1501 Tribunal Jeunesse 1502 Autre tribunal:…….. 1503 Service de probation 1504 Médiation pénale 1505 Altern. Détent. Prév. 1506 Com. Libér. Condit. 1507 Police 1508 Prison / SPS 1509 Avocat 1510 Com.défense sociale 1590 Autre:……	**1900 Handicap** 1901 B.R. AWIPH 1902 Serv. résidentiel 1903 Serv. d'acc. de jour 1904 Serv. d'accomp. 1905 Serv.d'aide précoce 1906 CRF non psy 1990 Autre:…… **2000 Santé mentale ambulatoire** 2001 « psy » privé 2002 autre SSM 2003 CRF psy 2004 Consult. psy en hôpital 2005 CPF 2006 SOS 2007 serv. Toxicomanie 2008 Un membre de l'équipe 2090 Autre:……	**2200 Milieu professionnel** 2201 Service social 2202 Médecin du travail 2203 Supérieur hiérarch. 2204 Collègue 2205 Ressourc. hum. 2290 Autre **9000 Autre** 9091…… 9092…… 9093…… 9094……
1200 Services sociaux 1201 CPAS 1202 Service AF.AS 1203 Mutuelle 1204 Maisons d'accueil 1205 Services sociaux de prévention : …… 1206 Centre d'accueil pour demandeurs d'asile 1207 Autre interculturel 1290 Autre :……..			
1300 Serv. petite enfance 1301 ONE (TMS, …) 1302 Gardienned'enfants 1303 Crèches 1304 Pouponnières 1305 Maisons maternelles 1390 Autre :……			

FIGURE 7.28 – Type de réseaux professionels

Codes R

Lecture des données

```
graphics.off()
library(lattice)
donnee<-read.delim(file="C:\\mémoire\\AdultesClasse.csv",header=TRUE,sep=";",
na.string="DM")
attach(donnee)

total<-nrow(donnee) #prend le nombre de lignes (nombre de patients)
entetes<-names(donnee)
```

Descriptif

```
################################################
#Histogrammes de l'âge (échantillon + Wallonie)
################################################

#Age de toutes les données

#### calcul de l'âge
Age<- AnneeEnregistrement- AnneeNaissance
#### suppression des âges incohérents
Age<- Age[(Age>18)&(Age<100)]
#### calcul des données pour l'histogramme
info<-hist(Age,breaks=1*17.5:99.5)
#Age d'une partie des données (ici exemple pour les PropPEC = "Bilan/expert"
#### calcul de l'âge des patients de cette catégorie
Age<- AnneeEnregistrement[PropositiondePriseenCharge.1.Niveau=="Bilan/Expert"] -
AnneeNaissance[PropositiondePriseenCharge.1.Niveau=="Bilan/Expert"]
#### suppression des âges incohérents
AgeSpec<- Age[(Age>18)&(Age<100)]
#### calcul des données pour l'histogramme
infoSpec<-hist(AgeSpec,breaks=1*17.5:99.5)

# tracage des deux histogrammes dans la même figure:

par(mfrow=c(1,2))
barplot(infoSpec$density,ylim=c(0,0.06),names.arg=infoSpec$mids,main="Histogramme
 de l'age des propositions de \n prise en charge 'Bilan/Expert'",xlab="Age",
 ylab="Densité",col="grey",cex.names=1.7,cex.axis=1.7,cex.lab=1.7,cex.main=1.7)
```

```
barplot(info$density,ylim=c(0,0.06),names.arg=info$mids,main="Histogramme
 de l'age des nouveaux patients ",xlab="Age",ylab="Densité",col="grey",
 cex.names=1.7,cex.axis=1.7,cex.lab=1.7,cex.main=1.7)
########################################
# Génération de tous les histogrammes
########################################

for(i in 4:ncol(donnee)){

#recuperation des donnees de la variable:
variable<-donnee[,i]
#prend une plus grande fenetre que par défaut pour avoir la place pour les labels:
par(mar=c(10,5.1,5.1,1.1))
#pour créer la table de fréquences: (une des deux lignes)
tab<-rev(sort(table(variable))) # pour trier les batons par ordre décroissant
tab<-table(variable) # pour laisser les batons par ordre des labels
#création du graphique et définition de ces parametres:
limiteeny<-max(tab)+5000 #limite axe y
nomchamp=entetes[i] #nom de la variable
#creation du graphique:
crt <- barplot(tab,ylim=c(0,limiteeny),las=2,col="grey",main=paste(nomchamp,'\n N=',
sum(tab)),cex.names=1.5,cex.axis=1.5,cex.lab=1.5)
#calcul et ajout des pourcentages sur chaque bâton
labels<-as.character(round((tab/total)*100,1),cex=0.8)
text(crt, tab, paste(labels,"%") ,adj=c(0.5,-0.5),cex=1.3)
dev.new()
}

####################################################
# Les 3 prises en charge proposées ensemble
# (le code pour le reseauProfessionnel est semblable)
####################################################

#creation d'une variable avec toutes les prises en charges proposées
variable <- rbind( as.character(PropositiondePriseenCharge.1), as.character
 (PropositiondePriseenCharge.2), as.character(PropositiondePriseenCharge.3))
 par(mar=c(12,6,5.1,1.1)) #prend une plus grande fenetre graphique
tab<-rev(sort(table(variable))) # pour trier les batons par ordre décroissant
tab<-tab[-c(which(names(tab) =="ErreurEncod"))] # retire la partie "ErreurEncod"
limiteeny<-max(tab)+5000
nomchamp="Toutes les propositions de prise en charge"
crt <- barplot(tab,ylim=c(0,limiteeny),las=2,col="grey",main=paste(nomchamp,"\n",
 "Chaque colonne = Pourcentage de patients concernés"),cex.names=1.5,cex.axis=1.5,
 cex.lab=1.5,cex.main=1.5)
labels<-as.character(round((tab/(sum(tab)))*100,1),cex=0.8)
text(crt, tab, paste(labels,"%",sep='') ,adj=c(0.5,-0.5),cex=1.3)


############################
#   Codes ICD10
############################

#pour prendre les valeurs commencant par "F":
variable<-Code.1a[substr(as.character(Code.1a),1,1) == "F"]
```

```
#pour prendre seulement les 2 premiers caractères des valeurs commencant par F ou G:
variable<-substr(as.character(Code.1a[substr(as.character(Code.1a),1,1) == "F" |
  substr(as.character(Code.1a),1,1) == "Z"]),1,2)

#pour prendre seulement le premier caractère
variable<-substr(as.character(Code.1a),1,1)

##############################
# Codes ICD10 principaux
##############################

variable<-Code.1a[substr(as.character(Code.1a),1,1) != "0"]
tab<-rev(sort(table(variable)))
nombreCode=sum(tab)
tab <- subset(tab, tab/nombreCode >0.02)
par(mar=c(3,6,5.1,1.1)) #prend une plus grande fenetre graphique
limiteeny<-max(tab)+1000
crt <- barplot(tab,ylim=c(0,limiteeny),las=1,col="grey",main="Codes principaux avec
 un effectif supérieur à 2%",cex.names=1.5,cex.axis=1.5,cex.lab=1.5,cex.main=1.5)
#pourcentage parmi ceux ayant au moins 1 code:
labels<-as.character(round((tab/nombreCode)*100,1),cex=0.8)
text(crt, tab, paste(labels,"%",sep='') ,adj=c(0.5,-0.5),cex=1.3)


##############################
# COdes F ou Z complet principaux
##############################

#pour prendre les valeurs commencant par "F" ou "Z":
variable<-Code.1a[substr(as.character(Code.1a),1,1) == "Z"]
tab<-rev(sort(table(variable)))
nombreF=sum(tab)
tab <- subset(tab, tab >300)
par(mar=c(3,6,5.1,1.1)) #prend une plus grande fenetre graphique
limiteeny<-max(tab)+1000
nomchamp="Codes principaux de type Z"
crt <- barplot(tab,ylim=c(0,limiteeny),las=1,col="grey",main=paste(nomchamp,"\n
 dont les effectifs sont supérieures à 300","\n Pourcentage au sein des codes
 'Z'"),cex.names=1.5,cex.axis=1.5,cex.lab=1.5,cex.main=1.5)
#pourcentage parmi ceux ayant au moins 1 code:
labels<-as.character(round((tab/nombreF)*100,1),cex=0.8)
text(crt, tab, paste(labels,"%") ,adj=c(0.5,-0.5),cex=1.3)

##############################
# Codes NI F NI Z complet principaux
##############################

#pour prendre les valeurs ne commencant pas par "F" ou "Z":
variable<-Code.1a[substr(as.character(Code.1a),1,1) != "Z"]
variable<-variable[substr(as.character(variable),1,1) != "F"]
variable<-variable[substr(as.character(variable),1,1) != "0"]
tab<-rev(sort(table(variable)))
nombreF=sum(tab)
```

```
tab <- subset(tab, tab >50)
par(mar=c(3,6,5.1,1.1)) #prend une plus grande fenetre graphique
limiteeny<-max(tab)+150
nomchamp="Codes principaux de type différent de F et Z"
crt <- barplot(tab,ylim=c(0,limiteeny),las=1,col="grey",main=paste(nomchamp,"\n
 dont les effectifs sont supérieures à 50","\n Pourcentage au sein des codes
 différents de Z et F"),cex.names=1.5,cex.axis=1.5,cex.lab=1.5,cex.main=1.5)
#pourcentage parmi ceux ayant au moins 1 code
labels<-as.character(round((tab/nombreF)*100,1),cex=0.8)
text(crt, tab, paste(labels,"%") ,adj=c(0.5,-0.5),cex=1.3)



##################################
# Codes F, 2 caractères par place
##################################

par(mfrow=c(3,1))
for(i in 1:3){
if (i==1){
variable<-substr(Code.1a[substr(as.character(Code.1a),1,1) == "F"],2,3)}
else if (i==2){
variable<-substr(Code.2a[substr(as.character(Code.2a),1,1) == "F"],2,3)}
else {variable<-substr(Code.3a[substr(as.character(Code.3a),1,1) == "F"],2,3)}
variable <- replace(variable, which(as.character(variable) == 99),"A")
variable<-paste("F",substr(variable,1,1),sep="")
tab<-table(variable)
par(mar=c(3,6,5.1,1.1)) #prend une plus grande fenetre graphique
limiteeny<-max(tab)+1500
nomchamp=paste("Codes F",i,"ème place")
crt <- barplot(tab,ylim=c(0,limiteeny),las=1,col="grey",main=paste(nomchamp,"\n
 Pourcentage au sein des codes 'F'"),cex.names=1.5,cex.axis=1.5,cex.lab=1.5,
 cex.main=1.5)
#pourcentage parmi ceux ayant au moins 1 code:
labels<-as.character(round((tab/sum(tab))*100,1),cex=0.8)
text(crt, tab, paste(labels,"%") ,adj=c(0.5,-0.5),cex=1.3)
}


####################
# Codes F, 2 caractères ENSEMBLES
####################

#d'abord joindre les 3 colonnes en une seule variable:
variable <- rbind( as.character(Code.1a), as.character(Code.2a), as.character(Code.3a))
#ne prendre que ceux commencant par F et que les deux derniers caractères:
#(car il fallait une nouvelle classe (FA))
variable<-substr(variable[substr(as.character(variable),1,1) == "F"],2,3)
variable <- replace(variable, which(as.character(variable) == 99),"A")
variable<-paste("F",substr(variable,1,1),sep="")
tab<-rev(sort(table(variable)))
par(mar=c(3,6,5.1,1.1)) #prend une plus grande fenetre graphique
limiteeny<-max(tab)+1500
nomchamp="Tous les codes F ensembles"
```

```
crt <- barplot(tab,ylim=c(0,limiteeny),las=1,col="grey",main=paste(nomchamp,"\n
 Pourcentage au sein des codes 'F'"),cex.names=1.5,cex.axis=1.5,cex.lab=1.5,
 cex.main=1.5)
#pourcentage parmi ceux ayant au moins 1 code
labels<-as.character(round((tab/sum(tab))*100,1),cex=0.8)
text(crt, tab, paste(labels,"%") ,adj=c(0.5,-0.5),cex=1.3)


############################
# 3 graphiques, 1 par code ICD10 (que premiere lettre)
############################

par(mfrow=c(1,3))
for(i in 1:3){
if (i==1){
variable<-substr(Code.1a,1,1)
}
else if (i==2){
variable<-substr(Code.2a,1,1)}
else {variable<-substr(Code.3a,1,1)}
for (j in 1:total){
if(variable[j]!="Z" && variable[j]!="F" && variable[j]!="0")
variable[j]="Autre"
}
par(mar=c(4,6,5.1,1.1)) #prend une plus grande fenetre graphique
tab<-rev(sort(table(variable))) # pour trier les batons par ordre décroissant
nbrDM=tab[c("0")] #pour savoir le nombre de données manquantes
tab<-tab[-c(which(names(tab) =="0"))]
limiteeny<-max(tab)+5000
nomchamp=paste("Code",i,"a")
crt <- barplot(tab,ylim=c(0,limiteeny),las=1,col="grey",main=paste(nomchamp,"\n
 (N=",total-nbrDM,")"),cex.names=2.0,cex.axis=2.0,cex.lab=1.7,cex.main=2.0)
labels<-as.character(round((tab/(total-nbrDM))*100,1))
text(crt, tab, paste(labels,"%",sep='') ,adj=c(0.5,-0.5),cex=2)
}

################################
# 3 codes IDC10 rassemblés- premier caractère
################################

#d'abord joindre les 3 colonnes en une seule variable:
variable <- as.vector(rbind( as.character(substr(Code.1a,1,1) ), as.character(
 substr(Code.2a,1,1)), as.character(substr(Code.3a,1,1) )))
par(mar=c(4,6,5.1,1.1)) #prend une plus grande fenetre graphique
for (j in 1:(3*total)){
if(variable[j]!="Z" && variable[j]!="F" && variable[j]!="0"){
variable[j]="Autre"
}}
tab<-rev(sort(table(variable))) # pour trier les batons par ordre décroissant
temp<-substr(Code.1a,1,1)
nbrDM1=table(temp)[c("0")] #pour savoir le nombre de données manquantes
tab<-tab[-c(which(names(tab) =="0"))]
limiteeny<-max(tab)+5000
nomchamp=paste("Code",i,"a")
```

```
crt <- barplot(tab,ylim=c(0,limiteeny),las=1,col="grey",main=paste("Codes ICD10
 (lettre)","\n","Chaque colonne = Pourcentage de patients ayant cette lettre"),
 cex.names=2.0,cex.axis=2.0,cex.lab=1.7,cex.main=1.8)
labels<-as.character(round((tab/(total-nbrDM1))*100,1),cex=0.8)
text(crt, tab, paste(labels,"%") ,adj=c(0.5,-0.5),cex=1.7)



#####################################
#  Mode de vie spécifique pour "Familial"
#####################################

variable<-Modedevie[as.character(ModedevieNiveau) == "Familial"]
par(mar=c(9,5.1,5.1,1.1)) #prend une plus grande fenetre graphique
tab<-rev(sort(table(variable))) # pour trier les batons par ordre décroissant
#  tab<-table(variable) # pour laisser les batons par ordre des labels
tab <- subset(tab, tab >2)
totalfam<-sum(tab)
limiteeny<-max(tab)+2000
crt <- barplot(tab,ylim=c(0,limiteeny),las=2,col="grey",main="Mode de vie familial",cex.names=1.5
labels<-as.character(round((tab/totalfam)*100,1),cex=0.8)
text(crt, tab, paste(labels,"%") ,adj=c(0.5,-0.5),cex=1.3)
dev.new()


#####################################
# Motifs pour "ProbRel" et "ProbPerso"
#####################################
par(mfrow=c(1,2))
variable<-MotifsPremièreConsultation[as.character(MotifsNiveau) == "ProbPer"]
par(mar=c(10,5.1,5.1,1.1)) #prend une plus grande fenetre graphique
tab<-rev(sort(table(variable))) # pour trier les batons par ordre décroissant
#  tab<-table(variable) # pour laisser les batons par ordre des labels
tab <- subset(tab, tab >2)
totalfam<-sum(tab)
limiteeny<-max(tab)+2000
crt <- barplot(tab,ylim=c(0,limiteeny),las=2,col="grey",main="Motifs pour
 'problème personnel'",cex.names=1.5,cex.axis=1.5,cex.lab=1.5)
labels<-as.character(round((tab/totalfam)*100,1),cex=0.8)
text(crt, tab, paste(labels,"%",sep='') ,adj=c(0.5,-0.5),cex=1.3)

variable<-MotifsPremièreConsultation[as.character(MotifsNiveau) == "ProbRel"]
par(mar=c(10,5.1,5.1,1.1)) #prend une plus grande fenetre

tab<-rev(sort(table(variable))) # pour trier les batons par ordre décroissant
#  tab<-table(variable) # pour laisser les batons par ordre des labels
tab <- subset(tab, tab >2)
totalfam<-sum(tab)
limiteeny<-max(tab)+2000
crt <- barplot(tab,ylim=c(0,limiteeny),las=2,col="grey",main="Motifs pour 'problème
 relationnel'",cex.names=1.5,cex.axis=1.5,cex.lab=1.5)
labels<-as.character(round((tab/totalfam)*100,1),cex=0.8)
text(crt, tab, paste(labels,"%",sep='') ,adj=c(0.5,-0.5),cex=1.3)
```

```
dev.new()


#############################################################
#  Niveau scolarité secondaire-derniere année réussie et type
#############################################################
par(mfrow=c(2,1))
i=18
variable<-donnee[,i][substr(as.character(donnee[,i]),1,1) == "4"]
variable<-substr(variable,2,2)
par(mar=c(3,5.1,5.1,1.1)) #prend une plus grande fenetre
# tab<-rev(sort(table(variable))) # pour trier les batons par ordre décroissant
tab<-table(variable) # pour laisser les batons par ordre des labels
tab <- subset(tab, tab >1)
tot=sum(tab)
limiteeny<-max(tab)+2000
nomchamp=entetes[i]
crt <- barplot(tab,ylim=c(0,limiteeny),las=1,col="grey",main=paste(nomchamp,"
 pour les secondaires"),cex.names=1.5,cex.axis=1.5,cex.lab=1.5)
labels<-as.character(round((tab/tot)*100,1),cex=0.7)
text(crt, tab, paste(labels,"%") ,adj=c(0.5,-0.5),cex=1.2)

i=17
variable<-donnee[,i][substr(as.character(donnee[,i-1]),1,3) == "2Se"]
par(mar=c(3,5.1,5.1,1.1)) #prend une plus grande fenetre
tab<-rev(sort(table(variable))) # pour trier les batons par ordre décroissant
tab <- subset(tab, tab >1)
tot=sum(tab)
limiteeny<-max(tab)+2000
nomchamp=entetes[i]
crt <- barplot(tab,ylim=c(0,limiteeny),las=1,col="grey",main=paste(nomchamp,"
 pour les secondaires"),cex.names=1.5,cex.axis=1.5,cex.lab=1.5)
labels<-as.character(round((tab/tot)*100,1),cex=0.8)
text(crt, tab, paste(labels,"%") ,adj=c(0.5,-0.5),cex=1.3)
dev.new()

###################################
# regarder les 3 colonnes de ressources en mm tps
# (similaire pour les prises en charges)
###################################


i=62 #endroit de la premiere colonne de ressource
#d'abord joindre les 3 colonnes de ressources en une seule variable:
variable <- rbind( as.character( donnee[,i] ), as.character( donnee[,i+1] ),
 as.character( donnee[,i+2] ))
par(mar=c(8,5.9,5.1,1.1)) #prend une plus grande fenetre

tab<-rev(sort(table(variable))) # pour trier les batons par ordre décroissant
tab<-tab[-c(which(names(tab) =="DM"))] # retire la partie "DM"
limiteeny<-max(tab)+5000
nomchamp=entetes[i]
```

```
#pour faire le pourcentage au sein des ressources:
#crt <- barplot(tab,ylim=c(0,limiteeny),las=2,col="grey",main=paste("Ressources",
 "\n","Chaque colonne = Pourcentage de cette ressource"),cex.names=1.5,cex.axis=1.5,cex.lab=1.5,c
#labels<-as.character(round((tab/sum(tab))*100,1),cex=0.8)

#pour faire le pourcentage des personnes ayant accès à cette ressource:
tabtemp<-table(donnee[,i])
nbrDM1=tabtemp[c("DM")]   #nombre de personnes avec aucune ressource (donc pas en 1)
crt <- barplot(tab,ylim=c(0,limiteeny),las=2,col="grey",main=paste("Ressources","\n",
 "Chaque colonne = Pourcentage de patients ayant cette ressource")
 ,cex.names=1.5,cex.axis=1.5,cex.lab=1.5,cex.main=1.3)
labels<-as.character(round((tab/(total-nbrDM1))*100,1),cex=0.8)


text(crt, tab, paste(labels,"%") ,adj=c(0.5,-0.5),cex=1.3)

##################################
# regarder les 3 colonnes de ressources séparément
# (similaire pour les prises en charges)
##################################

par(mfrow=c(1,3))
for(i in 62:64){
variable<-donnee[,i]
par(mar=c(9,5.9,5.1,1.1)) #prend une plus grande fenetre
tab<-rev(sort(table(variable))) # pour trier les batons par ordre décroissant
#tab<-table(variable) # pour laisser les batons par ordre des labels
nbrDM=tab[c("DM")] #pour savoir le nombre de données manquantes
if (nbrDM>0){
tab<-tab[-c(which(names(tab) =="DM"))] # retire la partie "DM"
if(i==64){
limiteeny<-max(tab)+500}
else {limiteeny<-max(tab)+3000}
nomchamp=entetes[i]
#avec le pourcentage de DM et le nombre de DM:
#crt <- barplot(tab,ylim=c(0,limiteeny),las=2,col="grey",main=paste(nomchamp,
 "\n","DM:",nbrDM, "soit", round((nbrDM/total)*100,1),"% des patients"),cex.names=1.5,cex.axis=1.

#avec le nombre de données restantes:
crt <- barplot(tab,ylim=c(0,limiteeny),las=2,col="grey",main=paste(nomchamp,"\n",
 "N=",total-nbrDM),cex.names=1.7,cex.axis=1.5,cex.lab=1.7,cex.main=1.5)
labels<-as.character(round((tab/(total-nbrDM))*100,1),cex=0.8)
text(crt, tab, paste(labels,"%",sep='') ,adj=c(0.5,-0.5),cex=1.6)
}
}


#######################################
#  Mode de vie spécifique pour "SansSourceRevenu"
# (similaire pour les catégories professionnelles)
#######################################
dev.new()
variable<-ModedevieNiveau[as.character(Source.Principale.de.Revenus) == "SansRevenu"]
par(mar=c(10,5.1,5.1,1.1)) #prend une plus grande fenetre
```

```
tab<-rev(sort(table(variable))) # pour trier les batons par ordre décroissant
tab <- subset(tab, tab >2)
totalfam<-sum(tab)
limiteeny<-max(tab)+1000
crt <- barplot(tab,ylim=c(0,limiteeny),las=2,col="grey",main="Mode de vie des patients
 \n Sans Source de Revenu",cex.names=1.5,cex.axis=1.5,cex.lab=1.5)
labels<-as.character(round((tab/totalfam)*100,1),cex=0.8)
text(crt, tab, paste(labels,"%") ,adj=c(0.5,-0.5),cex=1.3)



########################################
# Nature de la démarche par sexe
########################################


tab<-table(Sexe,NatureDemarche)
par(mfrow=c(1,2))
pie(tab[1,],cex=2,cex.main=2,radius = 1.3,main="Femmes",col=gray(seq(0,1,0.25)))
pie(tab[2,],cex=2,cex.main=2,radius = 1.3,main="Hommes",col=gray(seq(0,1,0.25)))

########################################
# mosaic plot (plusieurs méthodes)-bivariée
########################################

dev.new()
tab<-table(Sexe,NatureDemarche)
tab<-tab[,-c(1)]
tab<-prop.table(tab,2)
barplot(tab,legend.text = TRUE)


dev.new()
tab<-table(Sexe,NatureDemarche)

par(cex.main=1.5,cex.lab=1.5,mar=c(2.5,2.1,2.1,4.5))
mosaicplot(t(tab),las=2,col = hcl(c(240, 120,0)),cex=1.5,main="Mosaic plot")

par(las=2,mar=c(10.5,2.1,2.1,4.5),cex.lab=1.3,cex.axis=1.3)
spineplot(t(tab),col = hcl(c(240, 180,90,0)),cex=1.3,xlab="",ylab="")graphics.off()
```

Analyse des données manquantes

```
########################
# Pourcentage de données manquantes par SSM pour toutes les variables
########################

etab<-table(EtablissementXX)\\
#meme variable mais avec une colonne possédant le numéro de l'étab
etab2<-cbind(etab,names(etab))

for(i in 1:nrow(donnee)){
cas<-donnee[,i] #prend toutes les données de la variable i
```

```
variable<-table(EtablissementXX[cas=="DM"]) #prends les etablissement avec var(i)=DM
if(max(variable)>0){ #càd au moins une donnée manquante
#ajout du nom pour bien pouvoir joindre les deux variables:
var<-cbind(variable,names(variable))
tableau<-merge(var,etab2)
#convertir le tableau qui en en "list" en numérique:
X<-matrix(nrow=length(variable),ncol=4)
for (k in 1:nrow(tableau))
{
for (j in 1:length(tableau))
{
X[k,j]<-as.numeric(as.character(tableau[k,j]))
}
}
#calcul du pourcentage:
X[,4]<-round((X[,2]/X[,3])*100,2) #pourcentage de DM dans les établissements
colnames(X)<-c("Etab","variable","EffectifEtab","Pourcentage")
tab <- X[,4]
names(tab)<-as.character(X[,1]) #donne le nom de l'étab allant avec chaque pourcentage
tab<-rev(sort(tab))
print(tab)
par(mar=c(5,5.1,5.1,1.1))
limiteeny<-max(tab)+10
tab <- subset(tab, tab > 80 )
nomchamp=paste("Pourcentage de données manquantes dans la variable '",entetes[i],"'")
crt <- barplot(tab,ylim=c(0,limiteeny),las=1,col="grey",xlab="Numéro de
 l'établissement",main=nomchamp,cex.names=1.7,cex.axis=1.7,cex.lab=1.7)
text(crt, tab, paste(as.character(round(tab,1)),"%",sep='') ,adj=c(0.5,-0.5),cex=1.5)
dev.new()
}
}

################################
# définition des index des variables à analyser par la suite:
################################
vect<-c(16,19,22,23,62,30,56)

############################################
# Données manquantes par SSM - Histogramme et densité associée
############################################

etab<-table(EtablissementXX)
for(i in vect){
cas<-donnee[,i] #prend toutes les données de la variable i
variable<-table(EtablissementXX,cas) #prends les etablissement de ceux ou var(i)=DM
tab<-(variable[,c("0")]/etab)*100
par(mar=c(5,5.1,5.1,1.1))
nomchamp=paste("Histogramme des pourcentages de DM de chaque SSM \n dans la variable '",
 entetes[i],"'")
par(mfrow=c(1,2))
hist(tab,breaks=4,,main=nomchamp, xlab="Pourcentage de DM par SSM",cex.axis=1.5,cex.lab=1.5)
crt <- plot(density.default(tab,adjust=1),lwd=3,col="grey", main="Densité associée",
 xlab="Pourcentage de DM par SSM",cex.axis=1.5,cex.lab=1.5)
dev.new()
```

```
}

##########################################
# Données manquantes par Sexe - Histogramme (semblable pour l'année)
##########################################

sex<-table(Sexe)
par(mfrow=c(2,4))
par(mar=c(0.6,0.8,2,0.4))
pourcent<-round((sex/sum(sex))*100,1)
pie(sex,cex=2,cex.main=2,radius = 1.3,main="Toutes les données",labels = c(paste("DM \n",
 pourcent[1],"%"),paste("F \n", pourcent[2],"%\n"),paste("\n M \n",
 pourcent[3],"%")),col=gray(seq(0,1,0.25)))
for(i in vect){
par(mar=c(0.6,0.8,4,0.4))
cas<-donnee[,i] #prend toutes les données de la variable i
variable<-table(Sexe,cas) #prends les etablissement de ceux ou var(i)=DM
if (entetes[i]=="Code.1a"){
pourcent<-round((variable[,1]/sum(variable[,1]))*100,1)
pie(variable[,1],cex=2,cex.main=2,radius = 1.3,main=paste(entetes[i],"\n=DM"),labels
 = c(paste("DM \n", pourcent[1],"%"),paste("F \n", pourcent[2],"%\n"),paste("\n M \n",
 pourcent[3],"%")),col=gray(seq(0,1,0.25)))
}
else{
pourcent<-round((variable[,c("DM")]/sum(variable[,c("DM")]))*100,1)
pie(variable[,c("DM")],cex=2,cex.main=2,radius = 1.3,main=paste(entetes[i],"\n=DM"),
 labels = c(paste("DM \n", pourcent[1],"%"),paste("F \n", pourcent[2],"%\n"),paste("\n M
 \n", pourcent[3],"%")),col=gray(seq(0,1,0.25)))
}
}

##########################################
# Données RESTANTES par Sexe - Histogramme (semblable pour l'année)
##########################################

sex<-table(Sexe)
par(mfrow=c(2,4))
par(mar=c(0.6,0.8,2,0.4))
pourcent<-round((sex/sum(sex))*100,1)
pie(sex,cex=2,cex.main=2,radius = 1.3,main="Toutes les données",labels = c(paste("DM \n",
 pourcent[1],"%"),paste("F \n", pourcent[2],"%\n"),paste("\n M \n",
 pourcent[3],"%")),col=gray(seq(0,1,0.25)))
for(i in vect){
par(mar=c(0.6,0.8,4,0.4))
cas<-donnee[,i] #prend toutes les données de la variable i
variable<-table(Sexe,cas) #prends les etablissement de ceux ou var(i)=DM
if (entetes[i]=="Code.1a"){
ColonneDM<-variable[,1]
print(ColonneDM)
variableSansDM<-sex-ColonneDM
print(variableSansDM)
pourcent<-round((variableSansDM/sum(variableSansDM))*100,1)
pie(variableSansDM,cex=2,cex.main=2,radius = 1.3,main=paste(entetes[i],"\n sans DM"),
 labels = c(paste("DM \n", pourcent[1],"%"),paste("F \n", pourcent[2],"%\n"),
```

```
 paste("\n M \n", pourcent[3],"%")),col=gray(seq(0,1,0.25)))
}
else{
ColonneDM<-variable[,c("DM")]
print(ColonneDM)
variableSansDM<-sex-ColonneDM
print(variableSansDM)
pourcent<-round((variableSansDM/sum(variableSansDM))*100,1)
pie(variableSansDM,cex=2,cex.main=2,radius = 1.3,main=paste(entetes[i],"\n sans DM"),
 labels = c(paste("DM \n", pourcent[1],"%"),paste("F \n", pourcent[2],"%\n"),
 paste("\n M \n", pourcent[3],"%")),col=gray(seq(0,1,0.25)))
}
}
```

Comparaison à la Wallonie

```
########################
# Age
########################
Age<- AnneeEnregistrement - AnneeNaissance
Age<-Age[(Age>17)&(Age<100)]
info<-hist(Age,breaks=1*17.5:99.5)
curve(dnorm(x+17,mean=mean(Age,na.rm=TRUE),sd=sqrt(var(Age,na.rm=TRUE))),add=TRUE,col=2)


par(mfrow=c(1,2))
par(mar=c(5,5.1,5.1,1.1))
barplot(info$density,names.arg=info$mids,main="Histogramme de l'age des nouveaux
 patients \n adultes de 2008 à 2011",xlab="Age",ylab="Densité",col="grey",
 cex.names=1.7,cex.axis=1.7,cex.lab=1.7,cex.main=1.7)
curve(dnorm(x+17,mean=mean(Age,na.rm=TRUE),sd=sqrt(var(Age,na.rm=TRUE))),add=TRUE,col=2)
barplot(donneeWallAge[,2]/sum(donneeWallAge[,2]),ylim=c(0,0.03),names.arg=
 donneeWallAge[,1],xlab="Age",ylab="Densité",main="Histogramme de l'age des adultes \n
 en Wallonie en 2008",cex.names=1.7,cex.axis=1.7,cex.lab=1.7,cex.main=1.7)

plot(density.default(Age,na.rm=TRUE,adjust=1),main="Densité de l'âge",xlab="Âge",
 ylab="Densité",col=1,lwd=3)
curve(dnorm(x,mean=mean(Age,na.rm=TRUE),sd=sqrt(var(Age,na.rm=TRUE))),add=TRUE,col=2,lwd=3)
legend(80,0.025,c("Âge","Normale"),col=c(2,1),pch=15)

#graphique quantiles-quantiles de l'âge
qqnorm(Age);
qqline(Age, col = 2)
grid(lty=3)

#reproduction d'un vecteur des âges de la wallonie:
VectWall=rep(0,sum(donneeWallAge[,2])*1000)
dep=1
for (i in 18:99){
VectWall[dep:(dep+(donneeWallAge[i-17,2]*1000))]=i
dep=dep+(donneeWallAge[i-17,2])*1000
}
```

```
#graphique quantiles-quantiles entre notre échantillon et les données de la Wallonie
qqplot(Age,VectWall)

#test de student : l'échantillon a-t-il la mene moyenne que la Wallonie
t.test(Age,mu=mean(VectWall))


########################
# Sexe
########################

windows(record=TRUE, width=5, height=5)
par(pin=c(5,3))
par(mfrow=c(1,2))
par(mar=c(0,0,7,0))
tab<-table(Sexe)[-c(1)]
pourcent<-round((tab/sum(tab))*100,3)
pie(tab,col=gray(seq(0.5,1,0.25)),main="Répartition du sexe des nouveaux \n
 patients adultes de 2008 à 2011",labels=c(paste("F \n",pourcent[1],"% \n",seq=''),
 paste("\n M \n",pourcent[2],"%",seq='')),cex=1.7,cex.main=1.6)
SexeWall=c(5442557,5224309)
pourcent<-round((SexeWall/sum(SexeWall))*100,3)
pie(SexeWall,main="Répartition du sexe des adultes \n en Wallonie en 2008",
 labels=c(paste("F \n",pourcent[1],"% \n",seq=''),paste("\n M \n",pourcent[2],
 "%",seq='')),cex=1.7,col=gray(seq(0.5,1,0.25)),cex.main=1.6)

#test de proportion égales:
tabbis<-cbind(SexeWall,tab)
```

Dépendances par échantillonnage

```
#choisir une var1 et une var2

#var1="AnneeEnregistrement"
#var1="Sexe"
#var1="ModedevieNiveau"
#var1="AgeClasse"
#var1="Province"
#var1="CategorieProfessionnelle"
var1="Ressources.principale"


#var2="NatureDemarche"
#var2="TypeDossier"
#var2="OriginedelaDemarche"
#var2="DemandeduConsultant"
#var2="MotifsNiveau"
#var2="PropositiondePriseenCharge.1.Niveau"
var2="CodePrincClasse"

nbr<-100 #nombre d'echantillons
pvalinter<-mat.or.vec(1,nbr)
```

```
#boucle sur les variables

pval=0
ssdonnee<-donnee[,c(var1,var2)]

#boucle sur chaque echantillon
for(j in 1:nbr){
echant<-sample(1:dim(ssdonnee)[1],3875)
tbl<-table(ssdonnee[echant,2],ssdonnee[echant,1])
print(tbl)
if (var2=="NatureDemarche"){
tbl=tbl[-c(1),] #pour nature demarche car que des 0 dans "Autre"
}
if (var2=="OriginedelaDemarche"){ #grouper invalide et handicap et petite enfance dans autre
tbl[c(6),]=tbl[c(6),]+tbl[c(7),]
tbl[c(4),]=tbl[c(4),]+tbl[c(11),]+tbl[c(1),]
tbl=tbl[-c(1,7,11),]
}
if (var2=="MotifsNiveau"){
tbl[c(1),]=tbl[c(1),]+tbl[c(4),]
tbl=tbl[-c(4),]
}
if (var2=="PropositiondePriseenCharge.1.Niveau"){
tbl=tbl[-c(3,7,9),]
}
if (var2=="CodePrincClasse"){
tbl<-tbl[-c(2,7,10,11),]
}


if (var1=="ModedevieNiveau"){
tbl[,c(1)]=tbl[,c(1)]+ tbl[,c(5)]+tbl[,c(6)]+tbl[,c(2)]+tbl[,c(4)]
tbl[,c(2)]=tbl[,c(4)]+tbl[,c(2)]
tbl=tbl[,-c(4,5,6,2)]
}
if (var1=="NiveauScolariteAtteint"){
tbl[,c(5)]=tbl[,c(1)]+tbl[,c(4)]+tbl[,c(5)]
tbl=tbl[,-c(4,1)]
}
if (var1=="Province"){
tbl=tbl[,-c(4,1,2,5,6,9)]
}
if (var1=="AgeClasse"){
tbl<-tbl[,-c(4,5)] #sans les ages horslimites
}
if (var1=="CategorieProfessionnelle"){
tbl[,c(1)]<-tbl[,c(1)]+tbl[,c(2)]+tbl[,c(4)]+tbl[,c(5)]+tbl[,c(7)]
tbl<-tbl[,-c(2,4,5,7)]
}
if (var1=="Ressources.principale"){
tbl[,c(1)]=tbl[,c(1)]+tbl[,c(3)]+tbl[,c(4)]
tbl<-tbl[,-c(2,3,4)]
```

```
}


pvalinter[j]<-chisq.test(tbl)$p.value
pval<-pval + chisq.test(tbl)$p.value
}

# pour nature Demarche: pcq 0 autre tbl=tbl[,-c(1)]
# pour origine Demarche: tbl=tbl[,-c(7)]
print(var2)
print(tbl)
pv<-pval/nbr
print(pv)
#hist(pvalinter,breaks = 25)#
```

Classification

Classification intuitive

```
#########################
# création de la matrice intuitive:
#########################

nbrEchant<-3875

alea<-sample(1:total,nbrEchant)
donnee2<-donnee[alea,]
NbrLigne<-ncol(donnee)
dista=matrix(0, nbrEchant, nbrEchant)

for (l in c(1:nbrEchant)){
print(l)
for (k in c(l:nbrEchant)){
sum=0.0
for (j in c(1:NbrLigne)){
if (!(identical(donnee2[l,j],donnee2[k,j]))){
sum=sum+1
}
}
dista[l,k]<-sum
dista[k,l]<-dista[l,k]
}
}

index<-alea

distMat<-as.dist(array(dista))

#######################################
# création des dendrogrammes avec divers indices d'aggrégation:
#######################################

png("C:\\Users\\Media\\Dropbox\\Classif\\IntuitivSingleTTE.png")
```

```
hier<-hclust(distMat,"single")
plot(hier,hang=-1)
dev.off()


png("C:\\Users\\Media\\Dropbox\\Classif\\IntuitivCompleteTTE.png")
hier<-hclust(distMat,"complete")
plot(hier,hang=-1)
dev.off()

png("C:\\Users\\Media\\Dropbox\\Classif\\IntuitivCentroidTTE.png")
hier<-hclust(distMat,"centroid")
plot(hier,hang=-1)
dev.off()

png("C:\\Users\\Media\\Dropbox\\Classif\\IntuitivMedianTTE.png")
hier<-hclust(distMat,"median")
plot(hier,hang=-1)
dev.off()

png("C:\\Users\\Media\\Dropbox\\Classif\\IntuitivWardTTE.png")
hier<-hclust(distMat,"ward")
plot(hier,hang=-1)
dev.off()

################################
# récupération des classes:
################################

c2<-cutree(hier,2)
c3<-cutree(hier,3)
c4<-cutree(hier,4)

################################
# creation des classes:
################################

indexTrie<-sort(index)

ssdonnee<-donnee[index,]

################################
# visualisation dans la MCA:
################################

res.mcaClassif<-MCA(ssdonnee,na.method="NA")
plot(res.mcaClassif,choix="ind",axes=c(2,3),label=c("none"),invisible = c("var",
 "quali.sup"),col.ind=c3)


png("C:\\Users\\Media\\Dropbox\\Classif\\ClassifIntuitiv2Classes.png")
plot(res.mcaClassif$ind$coord[,c(1,2)],pch=as.character(c2),col=c2)
dev.off()
```

```
png("C:\\Users\\Media\\Dropbox\\Classif\\ClassifIntuitiv3Classes.png")
plot(res.mcaClassif$ind$coord[,c(1,2)],pch=as.character(c3),col=c3)
dev.off()

png("C:\\Users\\Media\\Dropbox\\Classif\\ClassifIntuitiv3ClassesBIS.png")
plot(res.mcaClassif$ind$coord[,c(1,3)],pch=as.character(c3),col=c3)
dev.off()

png("C:\\Users\\Media\\Dropbox\\Rfac\\Classif\\ClassifIntuitiv3ClassesTER.png")
plot(res.mcaClassif$ind$coord[,c(1,4)],pch=as.character(c3),col=c3)
dev.off()

png("C:\\Users\\Media\\Dropbox\\Classif\\ClassifIntuitiv3Classes5.png")
plot(res.mcaClassif$ind$coord[,c(1,5)],pch=as.character(c3),col=c3)
dev.off()

png("C:\\Users\\Media\\Dropbox\\Classif\\ClassifIntuitiv3Classes6.png")
plot(res.mcaClassif$ind$coord[,c(2,3)],pch=as.character(c3),col=c3)
dev.off()

####################################
# comparaison des variables au sein des classes
# ici exemple avec le code principal
####################################

donneeC1<-ssdonnee[c(c2==1),]
donneeC2<-ssdonnee[c(c2==2),]

#graphiques:

Classe1<-table(donneeC1$CodePrincClasse)
Classe1<-Classe1/sum(Classe1)
Classe2<-table(donneeC2$CodePrincClasse)
Classe2<-Classe2/sum(Classe2)
table<-rbind(Classe1,Classe2)
#pour province:
#table<-table[,-c(7)]

par(mar=c(10.5,3.1,2.1,1.5),cex.lab=1.3,cex.axis=1.3,cex=1.2,las=2)
barplot(table,beside=TRUE,ylim=c(0,max(table)+0.1),legend=TRUE,,args.legend=
 list(x="topleft"))

########################################
# enregistrement des données
########################################

write.csv(alea,file="IndexIntuiToutesVARSansDM.vsc")
write.csv(dista,file="MatriceIntuiToutesVARSansDM.csv")
```

Classification avec MCA et comparaisons

```
#install.packages("FactoMineR")
#install.packages("ca")
```

```
graphics.off()

library(lattice)
library(FactoMineR)
library(ca) #pour les MCA

#pour avoir le mm échantillon que intuitivement:
index<-read.csv(file="C:\\Users\\Media\\Dropbox\\IndexIntuiToutesVARSansDM.csv",
 header=TRUE,sep=",")
index<-index[,2]

###################################
# MCA
###################################

donnee2<-donnee[index,]

res.mca=MCA(donnee2)

###################################
# Classification
###################################

res.hcpc = HCPC(res.mca,nb.clust=6,metric="euclidian",method="ward")


dev.new()
plot(as.dendrogram(res.hcpc$call$t$tree))

dev.new()
plot(res.hcpc,axe=c(2,3),choice="map",draw.tree=FALSE,label=c("none"),cex.legend=1.3)

ssdonnee<-donnee[(row.names(res.hcpc$data.clust)),]
c2<-res.hcpc$data.clust[,13]

donneeC1<-ssdonnee[c(c2==1),]
donneeC2<-ssdonnee[c(c2==2),]

###################################
# Intuitive
###################################

data<-read.csv(file="C:\\Users\\Media\\Dropbox\\MatriceIntuiToutesVARSansDM.csv",
 header=TRUE,sep=",")


distance<-data[,2:3876]
distMat<-as.dist(array(distance))

hier<-hclust(distMat,"ward")
plot(hier,hang=-1)


c2<-cutree(hier,2)
```

```
ssdonnee<-donnee[index,]

donneeC1Intui<-ssdonnee[c(c2==1),]
donneeC2Intui<-ssdonnee[c(c2==2),]

###################################
# Comparaison
###################################

#graphiques:

dev.new()
par(mfrow=c(1,2),mar=c(10.5,3.1,2.1,1.5),cex.lab=1.3,cex.axis=1.3,cex=1.2,las=2)
Classe1<-table(donneeC1$Nationalite)
Classe1<-Classe1/sum(Classe1)
Classe2<-table(donneeC2$Nationalite)
Classe2<-Classe2/sum(Classe2)
table<-rbind(Classe1,Classe2)
ylimite=max(table)+0.3
barplot(table,beside=TRUE,ylim=c(0,ylimite),legend=TRUE,,args.legend=list(x="topright",
 bty="n"),main="MCA")

#pour province:
#table<-table[,-c(7)]

Classe1<-table(donneeC1Intui$Nationalite)
Classe1<-Classe1/sum(Classe1)
Classe2<-table(donneeC2Intui$Nationalite)
Classe2<-Classe2/sum(Classe2)
table<-rbind(Classe1,Classe2)
barplot(table,beside=TRUE,ylim=c(0,ylimite),legend=TRUE,,args.legend=list(x="topright",
 bty="n"),main="Intuitivement")
```