

Titre: Développement d'une typologie de ménages pour enrichir la
Title: prévision de la demande de transport

Auteur: Léa Fabre
Author:

Date: 2018

Type: Mémoire ou thèse / Dissertation or Thesis

Référence: Fabre, L. (2018). Développement d'une typologie de ménages pour enrichir la
Citation: prévision de la demande de transport [Mémoire de maîtrise, École Polytechnique de Montréal]. PolyPublie. <https://publications.polymtl.ca/3192/>

 **Document en libre accès dans PolyPublie**
Open Access document in PolyPublie

URL de PolyPublie: <https://publications.polymtl.ca/3192/>
PolyPublie URL:

**Directeurs de
recherche:** Catherine Morency
Advisors:

Programme: Génie civil
Program:

UNIVERSITÉ DE MONTRÉAL

DÉVELOPPEMENT D'UNE TYPOLOGIE DE MÉNAGES POUR ENRICHISSER LA
PRÉVISION DE LA DEMANDE DE TRANSPORT

LÉA FABRE

DÉPARTEMENT DES GÉNIES CIVIL, GÉOLOGIQUES ET DES MINES
ÉCOLE POLYTECHNIQUE DE MONTRÉAL

MÉMOIRE PRÉSENTÉ EN VUE DE L'OBTENTION
DU DIPLÔME DE MAÎTRISE ÈS SCIENCES APPLIQUÉES
(GÉNIE CIVIL)

AOÛT 2018

UNIVERSITÉ DE MONTRÉAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

Ce mémoire intitulé :

DÉVELOPPEMENT D'UNE TYPOLOGIE DE MÉNAGES POUR ENRICHIR LA
PRÉVISION DE LA DEMANDE DE TRANSPORT

présenté par : FABRE Léa

en vue de l'obtention du diplôme de : Maîtrise ès sciences appliquées

a été dûment accepté par le jury d'examen constitué de :

M. TRÉPANIÉ Martin, Ph. D., président

Mme MORENCY Catherine, Ph. D., membre et directrice de recherche

Mme SIOUI Louiselle, Ph. D., membre

REMERCIEMENTS

Si ce projet de recherche et ce double diplôme à l'École Polytechnique de Montréal, remplis de hauts et de bas, mais surtout de hauts, ont pu être menés à bien, c'est grâce à plusieurs personnes et organisations que je souhaite remercier.

Tout d'abord ma directrice de recherche, Catherine Morency, dont l'enthousiasme et l'énergie sont remarquables (et contagieux), pour m'avoir guidée tout au long de ce projet et pour m'avoir offert autant d'opportunités qui m'ont permis, durant ces deux ans, de découvrir et d'apprécier à chaque fois une nouvelle facette du monde des transports et de celui de la recherche. Je remercie également Hubert, Pierre-Léo et Jean-Simon pour leurs compétences et leurs précieux conseils, et la Chaire Mobilité, qui, menée par Catherine, fait régner un courant d'énergie positive au troisième étage du bâtiment de l'École Polytechnique! Je ne peux pas parler de Polytechnique sans remercier mes camarades de bureau, mon voisin de bureau Mikaël pour les expressions québécoises entre autres, Gab pour tous les « dépannages R » et tout le reste, Éli pour le soutien, les discussions et les « lifts »! Élo, pour la compagnie le soir tard au bureau et le soutien entre Françaises, Laurence pour les cours, les mordus, mais aussi les « beignes et café »! Laurent, pour les blagues et les anecdotes de grand-père, Anaïs, Greg, Jérôme et tous ceux qui sont passés par le B.330, vous avez tous contribué, chacun à votre manière, à rendre ces deux années agréables et mémorables!

En dehors des murs de Poly, je tiens à remercier tout particulièrement mes colocs, Bastian, Félix, Loïc, Maxime pour tous les bons moments partagés et surtout Clarisse pour m'avoir supportée en toutes circonstances, chapeau! Un grand merci à toute la gang du Club de Ski de Fond de l'Université de Montréal, pour m'avoir permis de m'évader dans les moments durs et moins durs de cette maîtrise! Je remercie également mes amis en France et bien sûr ma famille, mes parents tout particulièrement pour qui ces deux ans n'ont pas toujours été faciles, pour leur confiance en moi et leur soutien indéfectible, même de loin.

Enfin, la réalisation de ce projet n'aurait pu être possible sans le soutien financier du Ministère des Transports, de la Mobilité Durable et de l'Électrification des Transports, merci aussi à l'équipe de modélisation pour leur disponibilité et leurs conseils tout au long du projet.

Enfin, merci à l'INSA de Lyon de m'avoir donnée l'opportunité de faire ce double diplôme à l'École Polytechnique de Montréal, extraordinaire aventure tant au point de vue universitaire qu'humain.

RÉSUMÉ

De nos jours, de nombreux changements démographiques affectent la plupart des grandes villes du globe. Par exemple, les changements générationnels impliquent une population stagnante et vieillissante, alors que la taille des ménages diminue grandement, conséquence de nombreux bouleversements socio-économiques. Ces tendances s'accompagnent inévitablement de modifications dans les comportements de mobilité des personnes. En ce sens, il est pertinent de s'interroger sur la manière de planifier la demande en transport. Dans la province de Québec, au Canada, celle-ci se base uniquement sur les caractéristiques individuelles des personnes. Or, il semble évident que la composition d'un ménage influence les comportements de mobilité des personnes qui le composent. Ce projet de recherche répond donc à la problématique de prise en compte de l'attribut ménage dans l'identification des grandes tendances et hypothèses relatives à la mobilité des personnes. Les objectifs sont tout d'abord d'identifier les variables de caractéristiques de la personne et du ménage ayant une réelle influence sur les déplacements individuels. Ensuite, ils consistent en la création d'une typologie de ménages pertinente et transférable dans le temps, ainsi qu'à établir un modèle probabiliste permettant d'affecter les personnes aux types de ménages créés.

Dans la Grande Région de Montréal, la prévision de la demande en transports se fait à partir des données des enquêtes Origine-Destination (OD), réalisées tous les 5 ans. Les enquêtes des années 1998 à 2013, provenant du Ministère des Transports, de la Mobilité Durable et de l'Électrification des Transports (MTMDET), sont utilisées pour ce projet.

Pour répondre aux objectifs précédemment cités, une revue de littérature sur la démographie, les tendances de mobilité et les caractéristiques du ménage ayant une influence sur les déplacements des personnes est tout d'abord réalisée. Celle-ci permet de mettre en évidence les variables le plus souvent considérées lors de l'évaluation des impacts sur les comportements de mobilité. Elle permet également de conclure que la plupart des typologies de ménages jusqu'à présent effectuées ne permettent pas de classer toute la population et s'appuient souvent sur des catégories « classiques » de ménage.

Ensuite, les variables de la personne et du ménage sont analysées à l'aide d'arbres de décision et de régression afin de déterminer lesquelles ont le plus d'influence sur les déplacements individuels. Les arbres sont utilisés pour cinq indicateurs représentant les comportements de

déplacement de la personne : le fait d'être mobile ou non, le nombre de déplacements effectués dans une journée (sans les retours), la part de déplacements effectués en pointe matinale, la part de déplacements effectués en auto et finalement le fait de faire une chaîne complexe ou pas dans sa journée. Les arbres ne permettent pas de prédire correctement quelle valeur va prendre une observation donnée pour un indicateur de déplacement. Cependant, pour chaque indicateur et pour chaque année, les variables les plus déterminantes sont repérées. Pour l'ensemble des cinq indicateurs et des quatre années d'enquête, les variables qui ressortent comme les plus importantes sont l'âge moyen du ménage, l'amplitude maximale des âges dans le ménage, le nombre de personnes, la présence d'enfants de moins de 12 ans et celle de personnes âgées de 76 à 84 ans.

C'est avec ces variables que va être constituée la typologie de ménages à l'aide d'un algorithme de fouille de données, celui des k-moyennes. Celle-ci est constituée à partir des quatre années d'enquête rassemblées pour valider la transférabilité temporelle. Huit types de ménages sont créés. Les « jeunes couples, colocs ou célibataires », qui sont des ménages comprenant peu de personnes, pas d'enfant ni de personnes âgées et dont l'âge moyen et l'amplitude maximale des âges sont faibles. Les « cinquantenaires », dont les caractéristiques sont similaires à la catégorie précédente, si ce n'est que l'âge moyen est plus élevé. Les « retraités », pour lesquels on observe généralement la présence de 76-84 ans et dont l'âge moyen est plus élevé. Les « adultes et personnes âgées », plus rares, ce sont des ménages où l'amplitude maximale des âges est plus élevée, souvent constitués d'un adulte et d'une personne âgée comme son nom l'indique. Ensuite, trois groupes de familles : les « familles traditionnelles », les « très jeunes familles », les « jeunes familles », les familles traditionnelles étant composées d'enfants un peu plus âgés et de moins de personnes. L'âge moyen et l'amplitude maximale des âges sont plus bas chez les très jeunes familles que chez les jeunes familles. Enfin les « familles multigénérationnelles » sont des ménages plus atypiques, ayant un nombre de personnes plus élevé, et sont souvent constitués d'au moins un enfant, un adulte et une personne âgée.

Finalement, les variables entrantes dans le modèle de prévision du MTMDET sont utilisées pour constituer le modèle d'affectation qui permet de faire le lien entre les personnes de l'enquête OD et leurs comportements de déplacement suivant les attributs du ménage auquel elles appartiennent. Le modèle obtenu ne permet malheureusement pas d'obtenir les détails

d'affectation de chaque personne à un type de ménage, mais propose une répartition de la population au sein de ces derniers.

Certaines améliorations pourraient encore être apportées, particulièrement en ce qui concerne le modèle d'affectation (prise en compte de davantage de variables comme le statut par exemple). Ces perspectives de recherches, ainsi que les limitations rencontrées au cours du projet sont présentées à la fin de ce papier.

Le présent mémoire permet donc de mieux comprendre les interactions entre la structure des ménages et les comportements de mobilité des personnes qui les composent, ainsi que les effets relatifs à l'âge et au cycle de vie des personnes et de leur entourage sur leurs déplacements. L'étude de ces paramètres contribue à l'amélioration de la prévision de la demande en transport dans la grande région de Montréal et pourra bien sûr être appliquée à d'autres villes dans le futur. L'ensemble de ce travail s'insère dans un projet plus global qui vise à synthétiser la population et sa mobilité en prenant en compte les caractéristiques du ménage et les systèmes d'activités.

ABSTRACT

Nowadays, several demography changes are affecting most of the cities in the world. As an example, the population is getting older and the rate of growth appears to have come to a standstill. Also, household size diminishes due to several socio-economic changes. These trends inevitably go along with mobility behaviour adjustments. It is thus relevant to question the actual transport demand planification. In the Québec province, in Canada, it is uniquely based on individual characteristics and does not take into account the household characteristics. However, it seems clear that household attributes have an influence on the mobility behaviour of the people who belong to it. This research project answers the following problem: how to take into account the household attributes in the identification of the trends and hypotheses on the individual mobility. The objectives are: first, to identify the personal and household characteristics which truly have an impact on individual mobility; second, to create a household typology, which would be relevant and temporally transferable; third, to establish a probabilistic model which would affect people to a household type.

In the Great Montréal Area, the transport demand forecast is made thanks to the “Origine-Destination (OD)” surveys, conducted every five years. The surveys from 1998, 2003, 2008, and 2013 from the “Ministère des Transports, de la Mobilité Durable et de l'Électrification des Transports » (MTMDET) are used in this project.

In order to attain the objectives previously mentioned, a literature review is first conducted. It is about demography, mobility trends and household characteristics which have an influence on personal mobility behaviour. This review is a way to emphasize the variables most regarded when it comes to evaluating impacts on mobility behaviour. The review also enables one to notice that most of the household typologies already conducted are led by the “couple” notion and “classic” household types and do not classify all households.

Then the individual and household variables are analyzed thanks to decision and regression trees in order to determine those which have the biggest influence on individual trips. Five indicators are chosen to depict travel behaviour: being mobile or not, the number of trips in a day (not counting the returns home), the share of trips made during the morning peak, the share of trips made by car and doing a complex chain in the day or not. Trees do not enable to correctly predict the value one given observation will take for a given travel behaviour indicator. However, for

any indicator and any year of survey, the variables with the greatest impact are given. Hence, the selected variables at the end of this step are: mean age of the household, maximal amplitude of the ages in the household, number of people, presence of a child aged 12 or less, presence of a person aged between 76 and 84.

These variables are used to create the household typology thanks to a data-mining algorithm: the k-means. To build the typology, data of the four surveys are gathered in one database in order to get temporal transferability. Eight household types are created. The “young couples, flat mates or singles” which are households with few people, no children, no senior, a low mean age and low maximal amplitude. The “50th” which characteristics are similar to the previous category with a higher mean age. The “retired” for which we usually see presence of people aged 76-84 and a higher mean age. The “adults and older people”, less numerous, often composed of an adult and a senior as its name indicates. Then comes three family groups: the “traditional families”, the “very young families” and the “young families”. The traditional families are made with older children and fewer people. The mean age and the maximal amplitude of the ages in the very young families are lower than in the young families. Finally, the “Multigenerational families”, which are quite rare, with a higher number of people and often composed of at least a child, an adult and a senior.

In the end, the variables used for the prevision made by the MTMDET are used to build the affectation model which makes a link between the people of the OD survey and their travel behaviour depending on the household attributes. Unfortunately, the model obtained does not enable one to get the right household type affectation for each observation, but it does give a proportion of the population in every household types. Some improvements could be made, mainly for the affectation model (as an example, the use of additional variables such as the occupation). These research perspectives, as well as the limits encountered during the project are presented at the end of the paper.

Finally, this research project is a tool to better understand the interactions between the household structure and the mobility behaviour of the people in it, as well as the effects of age and life cycle of people and their relatives on individual trips. This study helps the improvement of travel demand forecast in the Great Area of Montréal and could, of course, be applied to another city in

the future. This work takes part in a bigger project which aims at synthesizing population and its mobility taking into account the household characteristics and the activity system.

TABLE DES MATIÈRES

REMERCIEMENTS	III
RÉSUMÉ.....	V
ABSTRACT	VIII
TABLE DES MATIÈRES	XI
LISTE DES TABLEAUX.....	XV
LISTE DES FIGURES.....	XVII
LISTE DES SIGLES ET ABRÉVIATIONS	XX
LISTE DES ANNEXES.....	XXI
CHAPITRE 1 INTRODUCTION.....	1
1.1 Mise en contexte.....	1
1.2 Problématique.....	1
1.3 Objectifs de recherche	3
1.4 Structure du mémoire	4
CHAPITRE 2 REVUE DE LITTÉRATURE	6
2.1 Évolution des tendances démographiques.....	7
2.1.1 Augmentation de la population	7
2.1.2 Vieillissement de la population	8
2.1.3 Évolution de la composition des ménages	9
2.1.4 Gradient spatial	10
2.2 Évolution des tendances de mobilité	11
2.2.1 Nombre de déplacements	11
2.2.2 Modes de transports	12
2.2.3 Types de déplacements.....	12

2.2.4	Différences entre les hommes et les femmes	13
2.3	Impacts des caractéristiques individuelles et de la structure du ménage sur la mobilité	14
2.3.1	Caractéristiques individuelles influençant les comportements de mobilité	14
2.3.2	Impacts du ménage sur les déplacements individuels	18
2.4	Typologies de ménage	21
2.5	Hypothèses tendanciennes de la demande de transport à Montréal	22
2.5.1	Les facteurs de pondération de l'enquête Origine-Destination	23
2.5.2	Reconstitution de la population et facteurs de pondération	24
2.5.3	Exemples de prévision de la mobilité par le MTMDET	26
CHAPITRE 3	MÉTHODOLOGIE	28
3.1	Méthodologie générale	28
3.2	Source de données	29
3.2.1	Description de l'enquête OD	30
3.2.2	Faits saillants des enquêtes OD	32
3.2.3	Travail effectué sur les données	35
3.3	Indicateurs	41
3.3.1	Indicateurs de mobilité	41
3.3.2	Variables explicatives	42
3.4	Arbres de décision	44
3.5	Méthodes de classification	49
3.6	Tests statistiques	51
3.6.1	Comparaison des moyennes	51
3.6.2	Comparaison des distributions	51
3.7	Modèle logit multinomial	53

CHAPITRE 4	SÉLECTION DES VARIABLES À INTÉGRER À LA TYPOLOGIE.....	55
4.1	Variables représentant les âges dans un ménage.....	55
4.2	Sélection des variables pertinentes pour expliquer les comportements de déplacement	60
4.2.1	Arbres de décision.....	61
4.2.2	Importance des variables et processus de sélection	71
4.3	Preuve que les variables sélectionnées ont bien une influence sur les déplacements des personnes.....	74
4.3.1	Combinaisons de types de ménages	75
4.3.2	Comparaison des déplacements de deux catégories de personnes	76
CHAPITRE 5	TYPLOGIES	80
5.1	Typologies de ménages année par année	81
5.1.1	Enquête 1998.....	82
5.1.2	Enquête 2003.....	84
5.1.3	Enquête 2008.....	85
5.1.4	Enquête 2013.....	87
5.1.5	Transférabilité temporelle	88
5.2	Typologie retenue.....	91
5.2.1	Typologie sur 4 années d'enquête	92
5.2.2	Différences de comportements de déplacement d'un type de ménage à un autre.....	96
5.2.3	Significativité des différences de comportement de déplacement : tests statistiques	102
5.3	Autres typologies testées.....	108
CHAPITRE 6	MODÈLE D'AFFECTATION DES PERSONNES À UN GROUPE DE MÉNAGES	110

6.1	Formation du modèle	111
6.2	Description du modèle	111
6.2.1	Résultats et performance du modèle	112
6.2.2	Utilisation du modèle	114
6.2.3	Évolution des répartitions.....	115
6.2.4	Prévisions pour 2021 et 2031	117
CHAPITRE 7	CONCLUSION	120
7.1	Synthèse de la recherche et contributions	120
7.2	Limitations	122
7.3	Recherches futures et perspectives.....	123
CHAPITRE 8	BIBLIOGRAPHIE	125
ANNEXES		131

LISTE DES TABLEAUX

Tableau 2.1 : Évolution de la couverture de l'enquête OD (Agence Métropolitaine de Transport, 2010).....	24
Tableau 3.1 : Description des champs de données des enquêtes OD pertinents pour l'étude	31
Tableau 3.2 : Description des champs de la base de données de chaînes des enquêtes OD pertinents pour l'étude.....	32
Tableau 3.3 : Description des variables avant et après agrégation.....	36
Tableau 3.4 : Description des variables avant et après agrégation (suite)	37
Tableau 3.5 : Liste des attributs de la table finale	40
Tableau 3.6 : Liste des attributs de la table finale (suite).....	41
Tableau 3.7 : Matrice de corrélation des variables sélectionnées pour les arbres de décision.....	44
Tableau 3.8 : Importance des variables de l'arbre de décision : être non mobile ou pas (enquête 2013).....	49
Tableau 4.1 : Matrice de confusion de l'arbre "Être mobile ou pas", enquête OD 2013.....	63
Tableau 4.2 : Matrice de confusion de l'arbre "Faire 3 déplacements", enquête OD 2013.....	66
Tableau 4.3 : Matrice de confusion de l'arbre "Part de déplacements en pointe AM", enquête OD 2013	68
Tableau 4.4 : Matrice de confusion de l'arbre "Part de déplacements en automobile", enquête OD 2013	70
Tableau 4.5 : Importance des variables pour l'année 1998	72
Tableau 4.6 : Importance des variables pour l'année 2003	72
Tableau 4.7 : Importance des variables pour l'année 2008	73
Tableau 4.8 : Importance des variables pour l'année 2013	73
Tableau 4.9 : Propriétés des catégories de ménage les plus fréquentes	77
Tableau 4.10 : Caractéristiques du ménage des deux catégories de personnes.....	78

Tableau 5.1 : Résultats du partitionnement de données pour l'enquête OD 1998.....	82
Tableau 5.2 : Résultats du partitionnement de données pour l'enquête OD 2003.....	84
Tableau 5.3 : Résultats du partitionnement de données pour l'enquête OD 2008.....	86
Tableau 5.4 : Résultats du partitionnement de données pour l'enquête OD 2013.....	87
Tableau 5.5 : Résultats du partitionnement de données sur l'ensemble des données (1998 à 2013)	92
Tableau 5.6 : Valeurs-p pour le test de Tukey sur la mobilité	103
Tableau 5.7 : Valeurs-p pour le test de Tukey sur le nombre de déplacements.....	104
Tableau 5.8 : Valeurs-p pour le test de Tukey sur la part de déplacements en pointe matinale .	104
Tableau 5.9 : Valeurs-p pour le test de Tukey sur la part de déplacements en automobile	105
Tableau 5.10 : Valeurs-p pour le test de Tukey sur le booléen faire une chaîne complexe ou non	105
Tableau 5.11 : Valeurs p pour le test de Kolmogorov-Smirnov sur la mobilité	106
Tableau 5.12 : Valeurs p pour le test de Kolmogorov-Smirnov sur la mobilité (suite)	107
Tableau 6.1 : Risques relatifs du modèle logit multinomial	112
Tableau 6.2 : Matrice de confusion du modèle logit multinomial (simulation 1).....	114

LISTE DES FIGURES

Figure 2.1 : Prévision des déplacements à l'horizon 2036 dans la GRM (MTMDET, 2017b)	26
Figure 2.2 : Prévisions des parts modales à l'horizon 2036 dans la GRM (MTMDET, 2017b) ...	27
Figure 3.1 : Schéma méthodologique	29
Figure 3.2 : Sous-ensembles de l'enquête OD	30
Figure 3.3 : Pourcentage de personnes par catégorie d'âge suivant l'année d'enquête	33
Figure 3.4 : Pourcentage de ménages par taille suivant l'année d'enquête.....	33
Figure 3.5 : Parts modales par année (territoire d'enquête 1998)	34
Figure 3.6 : Nombre de déplacements par personne en fonction de la présence de personnes âgées dans le ménage (enquête 2003)	35
Figure 3.7 : Illustration d'un arbre de décision : être non mobile ou pas (enquête 2013)	47
Figure 3.8 : Matrice de confusion arbre de décision : être non mobile ou pas (enquête 2013)	48
Figure 3.9 : Exemple d'écart maximal entre deux fonctions de répartition.....	52
Figure 4.1 : Répartition des ménages suivant leur taille, leur âge moyen et l'amplitude maximale des âges dans le ménage.....	56
Figure 4.2 : Nombre de ménages en fonction de l'amplitude maximale des âges dans le ménage et de la taille de celui-ci.....	57
Figure 4.3 : Nombre de ménages en fonction de l'amplitude maximum des âges dans le ménage et de la taille de celui-ci (suite)	58
Figure 4.4 : Nombre de ménages en fonction de l'âge moyen du ménage et de la taille de celui-ci	59
Figure 4.5 : Nombre de ménage en fonction de l'âge moyen du ménage et de la taille de celui-ci (suite).....	60
Figure 4.6 : Position de la section suivante dans le schéma méthodologique.....	61
Figure 4.7 : Arbre de décision "Être mobile ou pas", enquête OD 2013	63
Figure 4.8 : Matrice de confusion de l'arbre "Être mobile ou pas", enquête OD 2013	64

Figure 4.9 : Arbre de décision "Nombre de déplacements", enquête OD 2013.....	65
Figure 4.10 : Arbre de décision "Faire 3 déplacements", enquête OD 2013	65
Figure 4.11 : Matrice de confusion de l'arbre "Faire 3 déplacements", enquête OD 2013	66
Figure 4.12 : Arbre de décision "Part de déplacements en pointe AM", enquête OD 2013	67
Figure 4.13 : Matrice de confusion de l'arbre "Part de déplacements en pointe AM", enquête OD 2013.....	68
Figure 4.14 : Arbre de décision "Part de déplacements en auto", enquête OD 2013.....	69
Figure 4.15 : Matrice de confusion de l'arbre "Part de déplacements en automobile", enquête OD 2013.....	70
Figure 4.16 : Arbre de décision "Faire une chaîne complexe ou pas", enquête OD 2013	71
Figure 4.17 : Position de la section suivante dans le schéma méthodologique.....	75
Figure 4.18 : 15 principales catégories de ménage observées pour les hommes de 25 à 34 ans ..	76
Figure 4.19 : Nombre de déplacements par personne du groupe d'âge sélectionné suivant la catégorie de ménage (enquête 2013)	78
Figure 4.20 : Part de déplacements en auto par personne suivant la catégorie de ménage (enquête 2013).....	79
Figure 5.1 : Position de la section suivante dans le schéma méthodologique.....	80
Figure 5.2 : Appartenance au type de ménage suivant l'âge et le genre (enquête 1998).....	83
Figure 5.3 : Appartenance au type de ménage suivant l'âge et le genre (enquête 2003).....	85
Figure 5.4 : Appartenance au type de ménage suivant l'âge et le genre (enquête 2008).....	86
Figure 5.5 : Appartenance au type de ménage suivant l'âge et le genre (enquête 2013).....	88
Figure 5.6 : Répartition des groupes suivant l'âge moyen et l'amplitude max des âges, enquêtes 1998 et 2003	89
Figure 5.7 : Répartition des groupes suivant l'âge moyen et l'amplitude max des âges, enquêtes 2003 et 2008	90

Figure 5.8 : Répartition des groupes suivant l'âge moyen et l'amplitude max des âges, enquêtes 2008 et 2013	91
Figure 5.9 : Répartition des personnes suivant l'amplitude maximale des âges et l'âge moyen dans leur ménage (1998 à 2013)	93
Figure 5.10 : Appartenance au type de ménage suivant l'âge et le genre (enquêtes 1998 à 2013)	94
Figure 5.11 : Appartenance aux types de ménages à travers le temps	95
Figure 5.12 : Appartenance aux types de ménages à travers le temps pour les 40-49 ans.....	96
Figure 5.13 : Appartenance aux types de ménages à travers le temps pour les 70-79 ans.....	96
Figure 5.14 : Position de la section suivante dans le schéma méthodologique.....	98
Figure 5.15 : Part de personnes mobiles et non mobiles dans les 8 types de ménage	98
Figure 5.16 : Nombre de déplacements (excluant les retours) par personne dans les 8 types de ménage	99
Figure 5.17 : Part de déplacements en pointe matinale dans les 8 types de ménage	100
Figure 5.18 : Part de déplacements en auto dans les 8 types de ménage	101
Figure 5.19 : Part de personnes faisant une chaîne complexe ou plus dans les 8 types de ménage	102
Figure 5.20 : Exemple de regroupement de données avec la normalisation "min-max"	109
Figure 6.1 : Position de la section suivante dans le schéma méthodologique.....	110
Figure 6.2 : Matrice de confusion du modèle logit multinomial (simulation 1)	113
Figure 6.3 : Répartition de la population dans les types de ménages prévue à la,ide du modèle	115
Figure 6.4 : Courbes de tendance de l'évolution des parts de chaque type de ménage	116
Figure 6.5 : Prévision de l'évolution de la répartition des types de ménages pour 2021	117
Figure 6.6 : Prévision de la répartition de la population dans les différents types de ménages en 2021 et 2031	118

LISTE DES SIGLES ET ABRÉVIATIONS

La liste des sigles et abréviations présente, dans l'ordre alphabétique, les sigles et abréviations utilisés dans le mémoire ou la thèse ainsi que leur signification. En voici quelques exemples :

AMT	Agence Métropolitaine de transport
AQTr	Association Québécoise des Transports
CMM	Communauté Métropolitaine de Montréal
GRM	Grande Région de Montréal
MTMDET	Ministère des Transports, de la Mobilité durable et de l'Électrification des transports
OD	Origine-Destination
RTL	Réseau de Transports de Longueuil
STM	Société de transport de Montréal
STL	Société de transport de Laval

LISTE DES ANNEXES

ANNEXE A : Création de l'ID personne et de l'ID ménage (exemple avec l'enquête OD 2013)	131
ANNEXE B : Création des BD personnes et ménages (exemple avec l'enquête OD 1998).....	132
ANNEXE C : Construction d'un arbre et de sa matrice de confusion (exemple avec 3 déplacements pour 2013)	140
ANNEXE D : Construction de la typologie (exemple avec la base de données comprenant les quatre années).....	142
ANNEXE E : Construction du modèle logit multinomial et de la matrice de confusion	147
ANNEXE F : Valeurs p pour le test de Kolmogorov-Smirnov sur le nombre de déplacements	157
ANNEXE G : Valeurs p pour le test de Kolmogorov-Smirnov sur la part de déplacements en pointe matinale	158
ANNEXE H : Valeurs p pour le test de Kolmogorov-Smirnov sur la part de déplacements en auto	159
ANNEXE I : Valeurs p pour le test de Kolmogorov-Smirnov sur les chaînes complexes.....	160
ANNEXE J : Matrices de confusion des 10 simulations du modèle d'affectation	161

CHAPITRE 1 INTRODUCTION

1.1 Mise en contexte

Les premières modélisations de la demande de transport en milieu urbain datent du début des années 1950, aux États-Unis. À ce moment, en période post-guerre et d'essor de l'industrie automobile, une des priorités est de permettre la circulation des marchandises et des personnes (Bonnell, 2002). Les années 1950 sont synonymes d'investissements massifs dans les infrastructures de transport. Ces derniers s'accompagnent d'une entrée dans l'ère moderne de la technicité, avec l'apparition de nouveaux procédés qui sont les prémices de la modélisation en transport (système de comptage, premières enquêtes Origine-Destination...). Après cela, la modélisation de la demande de transport est devenue populaire et a été transférée d'abord en Europe puis dans le reste du monde. Les raisons de cet engouement sont nombreuses. Tout d'abord, il y a une volonté d'assurer une meilleure adéquation entre les offres de transport et les besoins des personnes. Outre cela, plusieurs exigences socio-économiques (évolution rapide des tendances de mobilité, rôle du transport dans le développement économique et social, coût des investissements dans les infrastructures de transport...) contribuent au développement des procédures de planification et des outils de modélisation (Bonnell, 2002).

C'est dans cette foulée que se sont développés les enquêtes Origine-Destination et les premiers modèles de transport : modélisation de la prévision du trafic aux États-Unis au début des années 1950, puis le fameux modèle à quatre étapes au début des années 1960 (Masson, 1998).

Dans la Grande Région de Montréal (GRM) et au Québec plus généralement, les enquêtes OD sont apparues en 1970. Depuis, la modélisation de la demande de transport a fortement été améliorée, mais celle-ci présente toujours certaines lacunes. En effet, il semble que certaines composantes ne soient pas intégrées au développement des scénarios tendanciels. Ce mémoire tente donc de participer au processus d'amélioration de l'approche de modélisation actuelle.

1.2 Problématique

En planification des systèmes de transport, la prévision de la demande de transport est primordiale. Pour prévoir les besoins de déplacement, il est nécessaire de comprendre les comportements de mobilité actuels ainsi que les principales tendances les affectant. Au Québec,

dans les grandes régions urbaines, c'est le service de la modélisation des systèmes de transport au Ministère des Transports, de la Mobilité durable et de l'Électrification des transports (MTMDET) qui a la mission de prévoir la demande. Le MTMDET s'appuie principalement sur les données recueillies lors des grandes enquêtes Origine-Destination (Morency, 2016) réalisées dans les grandes régions métropolitaines (tous les cinq ans à Montréal). Les prévisions se basent à la fois sur l'enquête OD la plus récente, mais également sur l'ensemble des enquêtes réalisées afin de prendre en compte les évolutions observées au fil du temps. C'est ainsi que sont formées par le MTMDET des hypothèses plausibles d'évolution des principaux facteurs explicatifs de la demande en transport pour les 25 prochaines années.

Cependant, des améliorations peuvent être apportées à cette approche afin de mieux prendre en compte certaines caractéristiques de la population et de ses déplacements, comme notamment les systèmes d'activités et les ménages.

Ce projet de recherche prend la suite d'un projet déjà réalisé à Polytechnique Montréal en collaboration avec le ministère : l'enrichissement du processus de modélisation de la demande de transport à l'aide des modèles « basés sur les activités » (Morency, C & Trépanier, 2013). Les modèles d'activités prônent le fait que ce ne sont pas les personnes qui dictent les besoins en déplacement, mais plutôt leur système d'activités. Ce projet avait pour but de transférer et d'expérimenter un modèle de génération de programme d'activités, développé à Toronto, à la région de Montréal. Le MTMDET pourrait faire évoluer ses pratiques de prévision de la demande de transport en s'appuyant sur ce projet et en mettant en place une modélisation basée activités pour la prévision de la demande de transport.

À cette fin, quelques améliorations pourraient être apportées à l'enquête OD et à son utilisation. Certaines pistes de bonification sont explorées par le projet de recherche R661.2 : « Synthétisation de la population et de sa mobilité : prise en compte des caractéristiques du ménage et des systèmes d'activités ». Ce projet, s'étalant sur trois ans, doit répondre à deux problématiques. Premièrement, l'intérêt de prendre en compte la structure des ménages dans les hypothèses tendanciellées formulées pour prévoir la demande en transport. En effet, le modèle actuel s'appuie uniquement sur les prévisions démographiques auxquelles des hypothèses de « taux de chef », c'est-à-dire la probabilité d'un individu d'être l'ancrage du ménage, sont ajoutées. Ces hypothèses permettent de dériver des projections de ménages, mais elles ne

permettent aucunement de construire explicitement les futurs ménages. Or, la structure des ménages est déterminante des comportements de déplacement individuels. Ainsi, elle a une influence non négligeable sur la prévision des comportements de mobilité. De plus, l'objet ménage est d'autant plus pertinent que les choix résidentiels et de motorisation, ainsi que celui du type et de la localisation du logement, dépendent de ce dernier. Les décisions à ces propos sont souvent prises au niveau du ménage et non de l'individu. Ajouter cette composante à la prévision de la demande de transport exercée par le ministère permettrait donc de gagner en justesse et en précision. La deuxième problématique de ce projet de recherche concerne le facteur d'expansion utilisé pour compenser le fait que les prévisions de déplacements se basent sur un échantillon pondéré, c'est-à-dire que chaque personne enquêtée en représente plusieurs. Le but est de réduire l'impact de l'échantillonnage en synthétisant la population.

Le projet de maîtrise exposé dans ce mémoire porte sur la première des deux problématiques annoncées précédemment, à savoir la prise en compte de l'attribut ménage dans l'identification des grandes tendances et hypothèses relatives à la mobilité des personnes. Ainsi, ce mémoire tente de répondre à deux grandes questions de la modélisation de la demande de transport : quelles sont les caractéristiques d'un ménage qui ont une influence sur les choix de mobilité de chaque individu du ménage? Et comment prendre en compte la composition des ménages dans le processus de prévision de la demande?

1.3 Objectifs de recherche

L'objectif général du projet de recherche auquel s'intègrent ces travaux de maîtrise consiste à améliorer l'approche actuelle de prévision de la demande en transport de façon à mieux tenir compte des systèmes d'activités des ménages, ainsi qu'à mieux considérer les synergies entre les ménages et les comportements de mobilité des personnes.

Le but du présent travail, qui se concentre sur le lien entre la composition des ménages et les comportements de déplacement des individus qui les composent est de « *développer une typologie de ménages pertinente pour l'analyse et la prévision des comportements de mobilité et assurer l'allocation des personnes à ces types de ménages dans les exercices de prévisions* ».

Pour cela, plusieurs sous objectifs devront être atteints :

1. Identifier les variables de caractéristiques de la personne et du ménage, disponibles dans l'enquête OD, ayant une grande influence sur les déplacements individuels.
2. Déterminer, parmi ces dernières, lesquelles seraient pertinentes à intégrer dans la typologie de ménages.
3. Développer une typologie de ménages pertinente et transférable dans le temps.
4. Évaluer l'évolution de l'appartenance à un groupe, mais aussi l'évolution des comportements de déplacement au sein d'un groupe à travers le temps.
5. Comparer les comportements de déplacement des personnes appartenant aux différents types de ménages obtenus précédemment et vérifier la significativité des différences d'un groupe à un autre.
6. Établir un modèle probabiliste permettant d'affecter les personnes aux groupes de ménages établis (en ne se basant que sur leurs caractéristiques individuelles projetables) pour ainsi faire le lien entre l'individu d'une enquête OD et la prévision de ses déplacements en fonction du ménage auquel il appartient.

1.4 Structure du mémoire

Le présent mémoire compte sept chapitres. L'introduction permet d'expliquer les raisons qui ont initié ce projet de recherche et de cadrer les limites des travaux couverts au cours de cette maîtrise parmi les objectifs plus nombreux du projet de recherche mandaté par le MTMDET. Ce premier chapitre renseigne également sur les objectifs du projet.

Le deuxième chapitre consiste en une revue de littérature sur l'objet ménage et les comportements de mobilité de la personne. Il permet notamment de faire un état des lieux de ce qui a déjà été traité à ce sujet et d'introduire ce qui le sera au sein de ce mémoire.

Le troisième chapitre correspond à une description de la méthodologie utilisée tout au long du projet. Les données utilisées y sont décrites et les méthodes de catégorisation des ménages utilisées sont expliquées.

Le chapitre quatre a pour objectif de décrire la manière dont les caractéristiques du ménage ayant une influence sur les déplacements individuels sont sélectionnées, pour ensuite être prises en compte dans la confection des typologies de ménages.

Le chapitre cinq donne des détails sur les typologies de ménages effectuées, sur leur transférabilité temporelle et sur la comparaison des comportements de déplacement entre les groupes obtenus.

Le sixième chapitre propose un modèle d'affectation des personnes à un type de ménage et évalue ce dernier.

Enfin le chapitre sept fait une conclusion du mémoire, énumère les limites rencontrées dans ce travail et propose des perspectives d'application et de recherche en lien avec le présent sujet.

CHAPITRE 2 REVUE DE LITTÉRATURE

Améliorer la mobilité des personnes est, de nos jours, pour la plupart des villes, un objectif majeur. Afin de procurer aux citoyens une offre de transport plus importante et mieux adaptée à leurs besoins, il est nécessaire de bien connaître la demande et ses évolutions. La mobilité des personnes est, pour cela, très étudiée dans les grandes villes. De nombreuses analyses sur les caractéristiques des personnes et leurs déplacements sont effectuées afin de prévoir la future demande en transport.

Plusieurs éléments influent sur les comportements de mobilité d'une personne. Il faut bien sûr penser à son âge, à son état de santé, sa situation géographique, ses revenus et ceci n'est qu'une liste non exhaustive. Des facteurs « externes » influencent aussi les choix de mobilité, comme la météo, l'offre en transport en commun... Plus encore, même si cela n'est pas toujours pris en compte dans les analyses, chaque personne fait partie d'un ménage, et les caractéristiques de ce dernier ont, elles aussi, un impact sur les choix de mobilité personnels.

La mobilité des ménages est un sujet d'étude en expansion depuis le début des années 1990 et est devenu un thème réellement important à partir du milieu des années 2000 (Web of science, 2018). Avant d'aller plus loin dans l'étude de ce sujet, il convient de définir le sens du mot ménage. En effet, les interprétations du mot « ménage », de nos jours, peuvent être diverses. Le terme ménage désigne souvent l'ensemble des personnes dont une famille est composée. En ce sens, plusieurs ménages peuvent cohabiter dans une même maison. Un ménage peut également désigner une « unité élémentaire de population d'un point de vue économique » (Reverso). Enfin le ménage peut être associé au foyer, c'est-à-dire le logement familial. Cette définition correspond donc à « l'unité statistique élémentaire de population » vivant sous un même toit, incluant donc la ou les personnes qui « quels que soient les liens qui les unissent occupent un même logement à titre de résidence principale » (Larousse, 2018). Ainsi le mot ménage peut se rapporter à différents concepts, principalement suivant l'époque. Dans le présent travail, la dernière définition est préférée étant donné qu'elle est adoptée dans l'enquête OD.

Un ménage s'organise généralement autour d'un « leader », le chef de famille, bien que ce concept tende à disparaître avec la diminution du nombre de ménages « traditionnels » observée depuis les années 1980 (Roussel, 1986). Le ménage est une unité de mesure souvent utilisée en

micro-économie et les données sont souvent collectées à ce niveau dans la plupart des études statistiques (Statistique Canada, 2015).

Les différentes compositions de ménage que l'on peut retrouver dans une métropole sont quasiment infinies. Il est donc difficile d'établir des catégories de ménages qui pourraient représenter correctement l'ensemble de la population montréalaise. Plusieurs chercheurs ont déjà tenté de faire des classifications de ménages. Arbonville et Bonvalet (1992), Strathman, Dueker et Davis (1994), Jarvis (1999a), Srinivasan et Bhat (2005) peuvent notamment être cités. Cependant, leurs travaux prennent rarement en compte plus de deux caractéristiques du ménage et se basent souvent sur des structures « traditionnelles ». Or ces dernières, fortement présentes jusqu'aux années 2000, tendent à laisser place à des formes de ménage moins classiques (familles recomposées, ménages monoparentaux, personnes seules, colocations, ...).

Les objectifs de la revue de littérature suivante sont au nombre de trois. Premièrement, il s'agit de faire un état des lieux des évolutions sur les dernières années, tant sur le point de la démographie que sur celui de la mobilité. Ensuite, il s'agira de relever les liens entre la composition du ménage et la mobilité identifiés dans les recherches effectuées. Finalement, une attention plus particulière sera portée aux typologies de ménage développées auparavant par les chercheurs.

2.1 Évolution des tendances démographiques

De nombreuses évolutions démographiques, ainsi que dans la composition des ménages, ont été observées dans les 30 dernières années dans les grandes villes de pays industrialisés. En effet, en plus de l'augmentation de la population mondiale, des changements spécifiques aux générations ou groupes de personnes sont notés. La population vieillit du fait de l'augmentation de l'espérance de vie et de la baisse du taux de natalité; la taille des ménages diminue et d'un point de vue spatial, la répartition de la population dans les villes suit un mouvement spécifique depuis les années 90.

2.1.1 Augmentation de la population

Les changements démographiques rythment l'évolution de la planète Terre depuis des millénaires et ceux-ci sont souvent la conséquence de transformations économiques et sociales.

La population mondiale approche les 7,6 milliards à la mi-2017, ce qui signifie que la planète compte environ un milliard d'habitants de plus qu'il y a 12 ans. Selon McEvedy et Jones (1978), cela prendrait seulement 13 ans à la population mondiale pour augmenter d'un milliard entre 2011 et 2024 alors que la même augmentation de population a pris 124 ans entre 1803 et 1927. Le XX^e siècle a donc marqué la démographie mondiale et le terme « explosion démographique » est ainsi né (Săvoiu & Simăn, 2017). De même, la population mondiale pourrait atteindre huit milliards en 2023 et dix milliards en 2055 (United Nations, 2017). Cependant, cette évolution est actuellement en train de stagner, voire de diminuer. En effet, le taux de croissance de la population diminue depuis 1985 (Polidano, 2017). Aussi, depuis la fin du siècle précédent, de grandes inégalités d'évolution de la population sont apparues. 97% de la croissance de la population serait exclusivement due à la croissance des pays pauvres ou en développement (Săvoiu & Simăn, 2017). En Europe et dans les pays développés, on parle même déjà « d'implosion démographique ».

2.1.2 Vieillesse de la population

La croissance de la population s'accompagne de variations dans la distribution de celle-ci. En effet, avec le déclin de la fécondité et l'augmentation de l'espérance de vie, le monde fait face à un vieillissement de la population. L'espérance de vie augmente, et ce, aussi bien dans les pays développés que dans les pays en développement (Chan, Saito, & Robine, 2016). Selon United Nations (2017), des gains significatifs ont été observés au niveau de l'espérance de vie, pour toutes les régions du monde entre 2000 et 2015. D'ailleurs, en 2015, l'espérance de vie est de 60,2 ans en Afrique, 74,6 ans en Asie, 77,9 ans en Europe et 79,2 ans en Amérique du Nord alors qu'en 1975, elle était inférieure à 75 ans en Europe et en Amérique du Nord, de moins de 60 ans en Asie et n'atteignait même pas 50 ans en Afrique. Ainsi, le fait que l'espérance de vie atteindrait un jour sa limite a même été réfuté (Vallin & Meslé, 2009). Par opposition, la fécondité, elle, subit un déclin sans précédent depuis les années 2000. Entre 2010 et 2015, 46% de la population vivait dans des pays où le taux de natalité était en dessous de 2,1 enfants par femme, entre 2045 et 2050, il est prévu que 69% de la population vive dans ces pays (United Nations, 2017). Ainsi, avec la population vivant de plus en plus longtemps et les naissances s'amenuisant, la moyenne d'âge de la population mondiale est de plus en plus haute. Depuis 1980, où l'âge médian de la population mondiale était de 22,6 ans, ce dernier n'a pas arrêté

d'augmenter (Communauté métropolitaine de Montréal, 2015). Il est même attendu que d'ici 2050, les 60 ans et plus représentent un quart de la population dans toutes les régions du monde excepté l'Afrique.

2.1.3 Évolution de la composition des ménages

Ces changements de population impliquent des changements dans la composition des ménages. En effet, le vieillissement de la population a pour conséquence une augmentation des ménages de petite taille (personnes âgées vivant seule ou à deux). Il en est de même pour la baisse de la fécondité qui entraîne une augmentation des ménages constitués d'un couple sans enfant. En France, par exemple, le nombre de personnes par ménage a régulièrement diminué entre 1968 et 2006 (de 3,06 il est passé à 2,30) et la part des ménages composés d'une seule personne est passée de 20% en 1968 à 33% en 2006 (Toulemon, 2011). À Montréal, la taille moyenne d'un ménage de la Grande Région de Montréal a progressivement diminué pour passer de 2,49 en 1998 à 2,37 en 2013 sur des territoires comparables (Secrétariat à l'enquête Origine-Destination, 2013). Cette augmentation du nombre de petits ménages est aussi due aux séparations et aux divorces qui créent des ménages d'une seule personne ou bien monoparentaux (Arbonville & Bonvalet, 1992). La part des ménages monoparentaux en France métropolitaine a augmenté de 4,8% entre 1968 et 2006 (Toulemon, 2011). Ces dernières catégories de ménages prennent de l'ampleur face à la catégorie des ménages « classiques » (deux adultes et deux enfants) qui n'est plus majoritaire (Roussel, 1986). Aussi, la part des ménages de plus de quatre personnes est maintenant davantage constituée de ménages complexes que de familles nombreuses (Statistique Canada, 2017). De plus, au sein de ces ménages complexes, le nombre de ménages regroupant des « personnes isolées » non âgées (amis, colocataires...) a augmenté par opposition à celui des ménages dont les personnes isolées correspondent à des personnes âgées (troisième génération) qui lui, a largement diminué (Toulemon, 2011). Les modifications de structure des ménages concernent aussi le statut des personnes, de plus en plus de ménages à travailleurs multiples étant observés (Strathman et al., 1994). Cette tendance s'accompagne de changements dans les attributions des tâches au sein du foyer.

2.1.4 Gradient spatial

Depuis les années 1920, la structure des villes (seules les villes de plus de 25 000 habitants sont concernées ici) évolue. Jusqu'en 1950, on observe une augmentation progressive de la densité de population. Celle-ci est principalement due à l'augmentation constante de la population. En revanche, de 1950 à 2000, la densité de population baisse largement, cela à cause de l'expansion des villes (annexion de territoires) premièrement, puis à partir de 1970 à cause d'une baisse de la croissance de population. Par exemple, la densité moyenne sur 119 villes des États-Unis en 1950 est de 8876 habitants par mile carré (3427 habitants par km²). En 2000, la moyenne sur ces mêmes villes est de 5659 habitants par mile carrée (2185 habitants par km²) (Kim, S., 2007). Aussi, en se penchant sur la structure spatiale des aires métropolitaines, certaines tendances sont observables. Plus on s'éloigne du centre-ville, plus la densité de population diminue, cela de manière linéaire. Cependant, le coefficient de décroissance de cette évolution s'est fortement affaibli de 1890 à 1950 et il diminue encore de 1950 à 2000 (Kim, S., 2007). Ces évolutions sont entre autres dues à l'augmentation du coût du logement en centre-ville, à la baisse de celui du transport ou bien à d'autres changements économiques ou politiques, qui ont encouragé de nombreuses personnes à déménager aux périphéries des villes. Selon Gaube et Remesch (2013), ces changements auraient non seulement une influence sur la structure socio-économique des aires urbaines, mais en plus, celle-ci serait plus importante que leur influence sur la densité de population. Dans une étude faite sur la ville de Vienne en Europe, bien que la structure du ménage ne soit pas la seule composante mise à l'épreuve (le revenu du ménage a beaucoup d'influence), certaines tendances apparaissent clairement. Ainsi, les jeunes ménages vivent principalement dans les zones pauvres ou à prix moyen, tandis que les familles sont plus souvent observées dans les périphéries (Collin & Mongeau, 1992). Souvent, les ménages de jeunes ou d'une personne, soit les ménages de petite taille, sont les plus pauvres (Gaube & Remesch, 2013) et se retrouvent dans les zones de la ville les moins coûteuses. À l'inverse, les ménages de plus grande taille, dont le revenu est souvent plus élevé, se retrouvent en banlieue, là où la vie est un peu plus chère. Selon Hurtubia, Gallay et Bierlaire (2010) aussi, les attributs du ménage sont les déterminants des « besoins et restrictions en termes de logement ». Ainsi, un changement de taille du ménage, dû à une naissance par exemple, donne souvent lieu à une relocation du ménage. En plus du nombre de personnes, d'autres attributs du ménage sont souvent cités comme ayant une influence sur la localisation et le type de logement dans lesquels ces ménages

vivent, ce sont notamment le revenu, le nombre d'enfants, l'âge, le groupe ethnique, le nombre de travailleurs, etc. À Montréal par exemple, une étude montre que « la pauvreté et le vieillissement sont concentrés au centre de l'agglomération » (Collin & Mongeau, 1992). L'influence de la taille du ménage sur son éloignement du centre-ville est confirmée pour la Grande Région de Montréal dans le rapport sur l'enquête Origine-Destination de 2013 (Secrétariat à l'enquête Origine-Destination, 2013). La taille moyenne des ménages est bien plus basse sur l'île de Montréal (2,22 personnes) qu'en banlieue (2,51 personnes pour la couronne sud ou 2,61 pour Laval). Cependant, cette différence pourrait s'atténuer étant donné que la taille des ménages tend à diminuer, cela particulièrement en banlieue où le taux de croissance des ménages est nettement supérieur à celui de la population (Communauté métropolitaine de Montréal, 2015).

2.2 Évolution des tendances de mobilité

Parallèlement aux changements démographiques, des changements dans les comportements de mobilité sont également observés. En effet, en même temps que la population augmente, que les villes s'étendent et que l'économie se développe, les besoins en mobilité et la demande de transport suivent le même mouvement. Cette évolution des comportements de mobilité est bien sûr spécifique aux grandes villes étant donné que l'offre en mobilité n'est pas comparable dans les villes de taille plus réduite. Dans les agglomérations, le nombre de déplacements par personne a subi plusieurs variations depuis les années 1980. Il en est de même pour la part modale des transports en commun et la motorisation. Enfin, même si actuellement certaines différences entre les déplacements des hommes et ceux des femmes sont toujours notables, au cours des dernières décennies, l'écart entre les deux s'est nettement amoindri.

2.2.1 Nombre de déplacements

Le nombre moyen de déplacements par personne a légèrement évolué entre 1987 et 2008, et ce, de manière différente suivant les catégories de population. Pour les plus jeunes, cet indicateur tend à diminuer, alors que le phénomène inverse est observé chez les plus âgés. De même, les femmes vivant dans des ménages de deux personnes font plus de déplacements en 2008 qu'en 1987 à partir de l'âge de 45 ans, alors que chez les hommes vivant dans la même catégorie de ménage, ce changement n'apparaît qu'à partir de 65 ans (Désilets, 2012).

Aussi, une diminution de la non-mobilité est notée entre 1987 et 1998, puis le taux stagne jusqu'en 2008 (Morency, C & Trépanier, 2013). Ceci est probablement dû au fait que, parallèlement à la diminution du nombre de personnes de tous âges ne se déplaçant pas, l'augmentation du nombre de personnes âgées (qui ont plus tendance à être immobiles) contrebalance cet effet.

2.2.2 Modes de transports

Un des changements les plus importants dans la mobilité des personnes sur les trente dernières années concerne la part modale du transport en commun. Celle-ci est très variable, après une diminution entre 1987 et 1998, une augmentation de la part modale des transports en commun est observée entre 1998 et 2008. Entre 2003 et 2008, le nombre de déplacements effectués en auto sur l'heure de pointe du matin dans la région montréalaise avait même légèrement diminué au profit de celui des déplacements en transport en commun qui avait, lui, augmenté de 14%. Pourtant, en 2013, l'augmentation du nombre de déplacements effectués en transport en commun n'était plus que de 8% tandis que les déplacements effectués en automobile avaient augmenté de 15% (Secrétariat à l'enquête Origine-Destination, 2013). L'engouement pour les transports en commun observée en 2008 n'a donc pas gardé la même importance jusqu'en 2013. En ce qui concerne les déplacements en mode actif, ils ne cessent d'augmenter depuis 1998 (Secrétariat à l'enquête Origine-Destination, 2013).

Du côté de la motorisation par personne, la tendance est à l'augmentation au cours des dernières décennies. En Inde par exemple, jusqu'en 2012, le taux de motorisation dans les grandes villes croît plus que la population (Reddy & Balachandra, 2012). À Montréal aussi, le nombre d'automobiles par personne (âgées de 16 ans ou plus) n'a cessé d'augmenter entre 1987 et 2008 (Désilets, 2012), même si cette augmentation est de moins en moins rapide. En 2013, le parc automobile est encore en hausse de 11% par rapport à 2008 (Secrétariat à l'enquête Origine-Destination, 2013).

2.2.3 Types de déplacements

De 1998 à 2008, il est également possible d'observer une augmentation de la part des déplacements motif travail effectués avant 7h du matin, pour certains segments de population, chez les hommes comme chez les femmes. Ceci, accompagné d'une augmentation de la durée

des activités travail entre 1998 et 2008 (Morency, C & Trépanier, 2013), implique un décalage des heures de pointe (plus tôt le matin et plus tard le soir).

Les chaînes de déplacements effectuées par les personnes ont aussi évolué au cours des dernières décennies. Une baisse générale du nombre de chaînes pour toutes les catégories de population est observée, cette baisse est plus importante de 1987 à 1998, mais se poursuit jusqu'en 2008. Par opposition, de 1987 à 1998, la complexité des chaînes a, elle, augmenté (Morency, C & Trépanier, 2013). De 1998 à 2008, l'évolution de la complexité des chaînes suit celle du nombre de chaînes par personne et par jour.

2.2.4 Différences entre les hommes et les femmes

Depuis longtemps, les différences entre la mobilité des hommes et celle des femmes sont avérées dans les aires urbaines. Du fait de leur rôle dans la structure sociale (Miralles-Guasch, Melo, & Marquet, 2016), les déplacements journaliers des femmes diffèrent de ceux des hommes. Par exemple, les femmes ont plus tendance à utiliser les transports en commun ou les modes doux, elles conduisent également moins loin pour aller au travail (Crane, 2007). Aussi, pour les déplacements dont le motif est le travail toujours, les hommes effectuent ces derniers plus tôt que les femmes (la part des déplacements travail avant 7h est plus importante chez les hommes que chez les femmes (Morency, C & Trépanier, 2013), et ce de manière significative). Cependant, ces écarts tendent à s'atténuer avec le temps (Morency, C & Trépanier, 2013), en effet, les femmes ayant de plus en plus accès au transport motorisé privé, utilisent moins les transports en commun et font des déplacements de plus en plus longs. Pourtant, ces tendances ne sont pas vraies pour toutes les femmes et pour tous les types de déplacements. En ce qui concerne les distances parcourues pour les déplacements motif travail, les différences de comportement entre les hommes et les femmes se sont amenuisées pour les personnes vivant seules et pour celles vivant en couple avec des enfants, mais au contraire, elles sont plus grandes pour les personnes vivant seule avec des enfants et pour les personnes vivant en couple sans enfant (Crane, 2007).

2.3 Impacts des caractéristiques individuelles et de la structure du ménage sur la mobilité

2.3.1 Caractéristiques individuelles influençant les comportements de mobilité

Les variables influençant les choix de mobilité de la personne sont très nombreuses. Elles peuvent être caractéristiques de la personne seulement (caractéristiques individuelles) ou bien de paramètres extérieurs comme la situation géographique (offre de transport en commun, topologie, climat...) ou bien encore de l'entourage de la personne.

En ce qui concerne les variables des caractéristiques individuelles influençant les choix de mobilité d'une personne, elles peuvent être intrinsèques (âge, genre...) ou bien extrinsèques (comme l'occupation, l'accès à une automobile...). Ces variables peuvent avoir une influence sur les choix de mobilité parce que la personne se retrouve dans l'incapacité de suivre un comportement de mobilité (par exemple utiliser une automobile pour se rendre au travail lorsque la personne n'y a pas accès) ou bien parce que la personne, suivant ses caractéristiques personnelles, va se sentir plus à l'aise avec un comportement. Il peut sembler peu rigoureux de généraliser des comportements suivant des caractéristiques de la personne, surtout intrinsèques, mais de nombreuses tendances ont été relevées dans la littérature et prouvent que certaines caractéristiques personnelles ont bien un impact sur les choix de mobilité. Celles-ci sont évidemment trop nombreuses pour toutes être citées, mais quelques-unes sont décrites ci-dessous.

2.3.1.1 Âge

Comme cela a été dit précédemment dans cette revue de littérature, la population mondiale vieillit et la part des 65 ans et plus dans les pays développés ne cesse d'augmenter. Bien que les personnes tendent à garder aussi longtemps que possible le comportement de déplacement qu'ils ont eu l'habitude de suivre (Alsnih & Hensher, 2003), les autorités responsables de la provision en moyens de transport ne peuvent ignorer le vieillissement de la population, car les besoins et les tendances de mobilité des personnes âgées ne sont bel et bien pas les mêmes que pour le reste

de la population. Premièrement, ces dernières se déplacent moins. À partir de l'âge de 80 ans, la réduction du nombre de déplacements devient évidente (Alsnih & Hensher, 2003).

Deuxièmement, en ce qui concerne les modes de transport utilisés, deux tendances contradictoires sont observées. Les personnes de 75 ans et plus utilisent davantage la voiture pour effectuer leurs déplacements que les transports en commun ou que la marche à pied. Ceci est particulièrement vrai en Europe, aux États-Unis et en Australie (Alsnih & Hensher, 2003). Cependant, il y a également une hausse de l'utilisation des transports en commun lorsqu'ils ne sont plus capables de conduire.

De manière générale, les personnes âgées comptent beaucoup sur leur entourage pour subvenir à leurs besoins en transport. Ceci est confirmé par une étude menée avec les données de la ville de Portland aux États-Unis (Hildebrand, 2003). Celle-ci propose de faire un portrait de la mobilité des personnes âgées. Il en ressort que la majorité des personnes âgées sont peu mobiles. Tout de même, une petite partie des personnes âgées est plus mobile que la moyenne (car ces personnes ont plus de temps pour faire diverses activités). Un grand nombre de personnes âgées dans cette étude sont très dépendantes de la voiture et souvent, elles ne sont pas conductrices et/ou le taux d'occupation de la voiture est élevé, c'est-à-dire que les personnes âgées sont souvent autopassagers dans leurs déplacements. Ceci confirme qu'elles dépendent de leur entourage pour leurs déplacements. Enfin, l'étude montre que les déplacements des personnes âgées se font dans un périmètre restreint autour de leur domicile.

À une échelle plus restreinte, des différences de comportement sont observées chez les hommes d'un même groupe, les « adultes », suivant leur âge : les hommes plus âgés (50-55 ans) sont plus susceptibles d'effectuer des trajets plus longs pour se rendre au travail que leurs homologues plus jeunes (McQuaid & Chen, 2012).

En ce qui concerne les enfants, ceux-ci effectuent majoritairement leurs déplacements en voiture, accompagnés d'un de leurs proches, et ce depuis quelques décennies. Cela est principalement dû au fait que la plupart du temps l'environnement n'est pas adapté pour que les enfants se déplacent de manière autonome (danger de circulation ou insécurité). Aux États-Unis par exemple, la distance vers l'école, les dangers de circulation, la météo, la criminalité, ... sont des critères souvent rapportés comme dissuadant les parents de laisser leurs enfants aller à l'école en marchant ou bien à vélo (Brownson & Boehmer, 2004).

Ainsi, il est évident que les enfants en bas âge et les personnes âgées ont des comportements de déplacement réellement différents de ceux des adultes de 20 à 60 ans. Ceci est donc à prendre en compte dans la planification des transports.

2.3.1.2 Genre

Plusieurs études ont montré des différences entre les tendances de déplacements des hommes et des femmes. Ces distinctions sont souvent attribuées au fait que les hommes et les femmes occupent des places différentes au sein du ménage. Pour Martel Poliquin (2012), le genre est et restera encore longtemps un déterminant du comportement de mobilité, car de nombreuses variables susceptibles d'influencer la mobilité lui sont liées (par exemple, le revenu influe sur les comportements de mobilité et notre société doit toujours faire face à des iniquités salariales). Ainsi, les femmes font souvent des trajets plus courts, mais plus nombreux dans une même journée, elles utilisent davantage les transports publics et font plus de trajets dont le but est d'accompagner quelqu'un (Bernard, Seguin, Bussiere, & Polacchini, 2000). Les hommes ont aussi tendance à effectuer des trajets plus longs pour se rendre au travail que les femmes : les hommes sont 1,5 fois plus susceptibles que les femmes d'effectuer des trajets de trente minutes ou plus (McQuaid & Chen, 2012).

Ces différences reflètent le fait que les femmes font leur choix de mobilité sous un panel de contraintes différent de celui des hommes (Sandow, 2008), en particulier la répartition des déplacements associés à l'entretien du ménage (Agarwal, 2004). Il suffirait que ces différences de « responsabilités » entre les sexes ne disparaissent pas pour que les différences de comportement de déplacements persévèrent elles aussi, peu importe les changements économiques et professionnels qui pourraient arriver (Bernard et al., 2000). Pourtant, dans les quatre dernières décennies, plusieurs tendances sont apparues dans la plupart des pays industrialisés : premièrement, la part de femmes sur le marché du travail et même sur les postes à hautes responsabilités a augmenté, deuxièmement, les rôles au sein d'un ménage (entre les hommes et les femmes) ont changé, et enfin, des changements dans la structure des ménages sont apparus et la population a vieilli. Ces trois tendances contribuent au fait que les comportements de déplacement des femmes sont de plus en plus similaires à ceux des hommes et que la société actuelle tend vers une convergence des comportements de mobilité (Rosenbloom, 2004).

Le genre apparaît donc comme un facteur important pour déterminer les comportements de déplacement, mais la prise en compte d'autres variables sociales est nécessaire. Par exemple, selon l'origine ethnique, les différences d'un genre à l'autre peuvent être plus ou moins marquées.

2.3.1.3 Occupation

De plus en plus, la prévision de la demande en transport se base sur les activités effectuées par les personnes dans une journée, car celles-ci ont une influence non négligeable sur les décisions prises par les personnes en ce qui concerne leurs déplacements (Lee, Hickman, & Washington, 2007). Ainsi, l'occupation de l'individu ne peut être oubliée lorsqu'il s'agit de déterminer les variables ayant une influence sur les déplacements individuels. Aussi, Scott et Kanaroglou (2002) s'appuient sur trois types de ménages pour modéliser le nombre de déplacements pour des activités autres que travail : les couples sans travailleur, les couples avec un travailleur et les couples avec deux travailleurs. En effet, le travail semble avoir un impact sur les activités autres que travail et donc sur les trajets pour s'y rendre, du moins le mode utilisé et le nombre d'arrêts effectués (Bhat & Singh, 2000). Cependant, le plus gros impact de l'occupation des personnes sur leurs trajets semble concerner les chaînes de déplacement et leur complexité. Les travailleurs développent des chaînes autour de leur déplacement pour le travail dans le but de satisfaire le maximum de leurs déplacements pour des activités autres que travail (Strathman et al., 1994). De plus, le statut retraité implique également des différences de comportement de déplacement, à la fois pour le mode choisi, mais aussi pour les motifs (plus de loisirs, visites chez le docteur...) (Hu, Wang, & Wang, 2013).

2.3.1.4 Autre

Il est facile de trouver encore d'autres variables ayant une influence sur les comportements de déplacement de la personne. Nombreuses sont d'ailleurs corrélées. Le revenu par exemple (également corrélé au statut) a de grosses influences sur les décisions de déplacement des personnes. Tout d'abord, le revenu a un impact sur le lieu de résidence. Il a en effet été reconnu que les personnes qui déménagent pour une zone de la ville différente (plus dense en l'occurrence) changent leurs comportements de déplacements (Buchanan & Barnett, 2006). Aussi, il est souvent mentionné que de faibles revenus impliquent un logement plus éloigné du centre, et donc des besoins de déplacements différents de ceux des personnes habitant en plein

centre-ville (souvent le cas de personnes ayant des revenus plus importants) qui peuvent, par exemple, se déplacer exclusivement à pied. Le revenu a également un impact sur l'accès à l'automobile. Ce dernier paramètre est déterminant du mode choisi pour la majorité des personnes : plus rares sont les personnes qui ont accès à une automobile, mais qui se déplacent en transports publics ou bien à pieds ou à vélo. Aussi, un accès à l'automobile faible implique un nombre de déplacements faible, surtout chez les adultes retraités ou ne travaillant pas (Bradley & Vovsha, 2005). Enfin, le motif du déplacement a un impact, principalement sur le mode choisi. Par exemple, pour magasiner, le mode préféré est la voiture (Buchanan & Barnett, 2006).

2.3.2 Impacts du ménage sur les déplacements individuels

Ces deux domaines d'évolutions sont bien sûr liés. Il a plusieurs fois été mentionné dans des travaux que la structure des ménages a un impact sur les comportements de mobilité individuels. Le nombre de personnes, leur genre, leur âge... tous les paramètres ayant une influence sur les déplacements individuels en ont également une sur les déplacements des autres personnes qui composent le foyer. En effet, les modèles d'activités et de déplacements des différents membres d'un même ménage sont interdépendants (Srinivasan & Bhat, 2005); (Kim, C. & Parent, 2016) et les choix de déplacement naissent souvent d'interactions avec les autres membres du ménage. Ainsi, les choix individuels de déplacements sont influencés par « l'existence, les opinions et les comportements des autres personnes » (Zhang, Kuwano, Lee, & Fujiwara, 2009). Dans un foyer, les décisions, y compris celles de déplacement, sont souvent prises de manière à ce que le niveau de satisfaction de tous les membres soit le plus haut possible. Les trajets communs sont également un facteur d'influence de la composition du ménage sur les déplacements individuels. Ceux-ci sont principalement observés pour les motifs magasinage et étude et, plus il y a de travailleurs dans un foyer, moins ils sont nombreux (Srinivasan & Bhat, 2005). Ces trajets groupés influencent un autre aspect du comportement de déplacement de la personne : le mode utilisé. En effet, les trajets groupés ont peu tendance à être effectués en transport en commun ou bien en transport actif.

2.3.2.1 Genre des autres membres du ménage

Outre la prise en compte des autres membres du ménage dans les choix individuels de déplacements, la structure du ménage a une réelle influence sur les déplacements à effectuer pour

chaque membre. Dans le cas des personnes vivant en couple par exemple, des différences entre les déplacements des hommes et des femmes ont souvent été relatées dans la littérature. Les femmes ont tendance à passer plus de temps pour les déplacements en lien avec l'entretien du ménage que les hommes et, au contraire, les temps de déplacement pour le travail sont plus importants chez les hommes que chez les femmes (Fan, 2017). Les femmes font aussi généralement plus de déplacements dans une journée que les hommes (Sweet & Kanaroglou, 2016) et leurs chaînes de déplacements diffèrent de celles effectuées par les hommes (type, complexité...). Les femmes sont également plus susceptibles d'inclure des passagers dans leurs déplacements (Scheiner & Holz-Rau, 2017). Selon Solá (2016), il existe un « contrat des genres » au sein d'un ménage, celui-ci peut être « traditionnel » (favoriser les activités des hommes aux dépens de celles de femmes), « équitable » (les activités des femmes sont autant favorisées que celles des hommes) et « mixte » (dans certains domaines les activités des hommes sont favorisées tandis que dans d'autres ce sont celles des femmes qui le sont). Ces contrats ont une influence sur plusieurs paramètres comme l'utilisation de la voiture, la prise en charge des enfants, etc., et donc, bien sûr, une influence sur les déplacements des hommes et des femmes. Cependant, pour Fan (2017), ces différences de comportement de déplacement entre les hommes et les femmes au sein d'un foyer dépendent aussi des origines ethniques (pour certaines communautés, les différences hommes/femmes sont plus présentes que pour d'autres) et elles sont susceptibles de changer dans le temps, surtout dans la société moderne actuelle, où de nombreux paramètres comme le mariage ou la cohabitation ne cessent d'évoluer.

2.3.2.2 Âge des autres membres du ménage

Aussi, plusieurs différences de comportement de déplacement ont pour cause la maternité davantage que le fait d'être une femme (Scheiner & Holz-Rau, 2017). Chez McQuaid et Chen (2012), pour la durée des trajets de motif travail par exemple, des différences entre les hommes et les femmes sont notées et, plus particulièrement, à l'intérieur de chaque groupe, des différences prévalent selon la présence d'enfant de moins de 19 ans. Pour Fan (2017), les différences de temps de trajet entre les hommes et les femmes pour le travail sont plus importantes en présence d'enfants. Ces écarts sont moins clairs lorsqu'il n'y a pas d'enfants dans le ménage. La présence d'enfants entraîne une répartition des déplacements en lien avec l'entretien du foyer plus inégalitaire (Ta, Kwan, Chai, & Liu, 2016) et elle implique de

nombreux trajets « contraints » pour les femmes, mais pas pour les hommes (Sweet & Kanaroglou, 2016). C'est le cas parce que les femmes ont plus tendance à prendre en charge les enfants et leurs activités. Selon une étude tirée de la *National Household Travel Survey* aux États-Unis, les personnes accompagnant les enfants pour leurs trajets sont les mères à 30% contre seulement 11% pour les pères (McDonald, 2008). Aussi, cette même étude relate que lorsque les enfants atteignent l'âge de conduire, la part de leurs déplacements effectués accompagnés de leur mère diminue tandis que la part des déplacements effectués accompagnés de leur père n'est pas vraiment impactée. Par conséquent, pour les femmes, avoir des enfants est synonyme de moins de trajets de type loisir (Bernardo, Paleti, Hoklas, & Bhat, 2015), mais aussi de plus de trajets plus courts (Sandow, 2008) et de chaînes plus complexes (Scheiner & Holz-Rau, 2017). La présence d'enfants dans un ménage exacerbe donc les différences de comportements entre les femmes et les hommes, mais plus généralement, elle impacte les déplacements de tous les membres du foyer. En moyenne la moitié des déplacements des enfants pour aller à l'école sont effectués accompagnés d'un membre de leur foyer, et un changement d'heure de début de l'activité scolaire impacterait les comportements de déplacement de tous les membres du ménage (McDonald, 2008). Aussi, plus la proportion d'enfants dans le ménage est grande, plus le nombre de déplacements par personne augmente (Bernard et al., 2000). C'est particulièrement le cas en présence d'enfants de plus de 6 ans, qui commencent à avoir des activités personnelles auxquelles ils ont besoin d'être amenés ou récupérés. Ce sont dans ce cas d'autres membres du foyer qui se chargent de ces déplacements (Bhat et al., 2013). « Le comportement des adultes vis-à-vis des déplacements est explicitement lié aux activités de leurs enfants (trajets pour l'école et pour les activités postsecondaires) » [Notre traduction] (Lee et al., 2007). Ceci implique donc des contraintes et un remaniement des déplacements de ces personnes en fonction des besoins des plus jeunes membres du ménage. La présence d'enfants dans un foyer aurait également un impact sur la distance des déplacements effectués par les membres du ménage, la présence d'enfants étant le deuxième critère d'influence sur la distance des déplacements (après l'âge de la personne) (McQuaid & Chen, 2012). L'âge du plus jeune enfant ferait également partie de ces critères. Un ménage comprenant des enfants en bas âge effectue, en moyenne, des déplacements plus courts (McBride, Davis, Lee, & Goulias, 2017). L'influence se fait également sentir sur la propension à effectuer des chaînes complexes (la présence d'enfants est synonyme de davantage de chaînes complexes). Les ménages monoparentaux avec

de jeunes enfants sont ceux qui effectuent les chaînes les plus complexes et en plus grand nombre (Bhat et al., 2013). Selon Hensher et Reyes (2000), la présence d'enfants est aussi synonyme de moins de chaînes simples impliquant les transports en commun pour se rendre au travail. Concernant les modes de déplacement, la présence d'enfants mineurs dans un ménage vient avec une utilisation de l'automobile plus accrue. En effet, l'utilisation d'autres modes de transport avec un enfant est plus compliquée puisqu'elle nécessite souvent plus d'organisation et même l'acquisition de matériel (casque pour le vélo par exemple) (Pooley et al., 2011). Les comportements de déplacement des membres d'un ménage sont donc impactés par la présence d'enfants « dépendants » principalement, mais des différences sont aussi visibles chez les adultes plus âgés, entre ceux qui ont eu des enfants (qui ne sont plus dépendants) et ceux qui n'en ont pas eu (McQuaid & Chen, 2012).

La présence de personnes âgées au sein d'un ménage aurait également une influence sur les déplacements d'un foyer. Celle-ci implique une diminution du nombre de déplacements. Pour un individu, le nombre de déplacements augmente avec l'âge jusqu'au groupe des 35-44 ans (McBride et al., 2017). Après cela, à partir de 55 ans, le nombre de déplacements par personne diminue avec l'âge (Lee et al., 2007). Aussi, un ménage composé de davantage de personnes âgées est un ménage qui se déplace moins. Certaines études montrent que le nombre de personnes dans un ménage a aussi une influence sur les déplacements de chacune d'entre elles. Par exemple, Shay et Khattak (2012) et Kato et Matsumoto (2009) montrent que les personnes appartenant à des ménages nombreux ont plus tendance à utiliser le transport en commun aux dépens de l'automobile. Ceci est dû au fait qu'au sein de ménages nombreux, les voitures doivent être partagées entre un plus grand nombre de personnes. Cependant, l'âge des membres du foyer impacte davantage les comportements de mobilité du ménage que le nombre de personnes qui y vivent (Théberge-Barrette, 2013).

2.4 Typologies de ménage

Ainsi, quelques travaux ont déjà tenté de regrouper les ménages suivant leur composition dans le but d'attribuer un type de ménage à chaque personne en fonction de ses caractéristiques individuelles ou bien de déceler des différences dans les comportements de déplacement suivant les types de ménage. Souvent ces regroupements se basent sur le nombre de personnes dans le ménage, le statut de celles-ci et la situation maritale du foyer (en couple avec enfants, en couple

sans enfant, ménage monoparental, personne seule, ...). Par exemple, Désilets (2012) utilise une typologie afin d'intégrer une variable « ménage » dans la prévision de la demande en transport. Cette typologie possède trois niveaux : en premier la taille du ménage, ensuite la présence d'enfants en fonction du nombre d'adultes et enfin le statut des individus (travailleur, étudiant ou autre). Vingt-deux types de ménages sont donc obtenus. Chez Bernard et al. (2000), la typologie se base sur la notion de couple. Ainsi, les différentes catégories de ménage retenues sont : couple, couple avec enfants mineurs, couple avec enfants majeurs, plusieurs adultes sans enfant, personne seule, personne seule avec enfants et des catégories « résiduelles ». Roussel (1986) lui, sépare les ménages en deux catégories : les « simples » (dont le noyau est familial) et les « complexes » (admettant des tiers). Une autre catégorisation basée sur la notion de couple est faite avec Srinivasan et Bhat (2005). Ici, quatre types de ménages sont considérés : les ménages composés d'un seul travailleur et ne comportant pas d'enfant, les ménages sans enfant, mais avec deux travailleurs, les ménages avec un enfant ou plus et un seul travailleur et les ménages avec un enfant ou plus et deux travailleurs. Cette dernière reste très limitée. De même, Strathman et al. (1994) n'utilise que le statut des personnes et la présence d'enfants et de personnes âgées. Enfin, Arbonville et Bonvalet (1992) prennent en compte davantage de critères. Quatre catégories de ménages sont considérées : les personnes seules, les ménages monoparentaux, les couples et les ménages « autres ». Pour chaque catégorie, le nombre de femmes et d'hommes est pris en compte, ainsi que le statut marital et la tranche d'âge (moins de 30 ans, 30 à 39 ans, 40 à 59 ans et 60 ans et plus). Pour les couples, le critère présence d'enfants est ajouté. Dans ces regroupements, la notion de couple prime souvent et les ménages représentés sont ceux dits « traditionnels ». Cependant, ceux-ci sont de moins en moins présents dans la population. Aussi, une catégorie « autres » est souvent ajoutée, cette dernière regroupe tous les ménages qui n'ont pas pu être classés et sera donc difficile à analyser, d'où le choix, dans la présente étude, d'établir une nouvelle typologie de ménages.

2.5 Hypothèses tendanciellles de la demande de transport à Montréal

À Montréal et au Québec, la planification des transports ne repose pas sur le modèle à 4 étapes, mais sur une approche de modélisation basée sur les enquêtes OD. Cette approche utilisée par le MTMDET a pour but de prévoir la demande en tirant « profit de toute la richesse des enquêtes

Origine-Destination québécoises » (MTMDET, 2017b). Elle consiste à reconstituer la population future par l'intermédiaire de facteurs de pondération. À cette reconstitution s'ajoutent des hypothèses sur les comportements de mobilité (issues d'analyses tendanciennes). C'est comme ça que sont obtenues des estimations de mobilité à plus ou moins long terme. Depuis 1993, la mise à jour de ces prévisions est effectuée tous les cinq ans. Celles-ci s'appuient principalement sur la précédente enquête OD réalisée et les enquêtes antérieures sont utilisées pour dégager certaines tendances. D'autres données sont également utilisées dans cette approche par exemple les données de Statistique Canada (MTMDET, 2017b). La plupart des facteurs pris en compte dans cette prévision de la demande concernent la population. Ainsi, l'évolution des déplacements est tirée de l'évolution de la population et correspond donc aux comportements de déplacement de celle-ci suivant ses caractéristiques.

2.5.1 Les facteurs de pondération de l'enquête Origine-Destination

L'enquête OD couvre 9 840 km² en 2013 soit 159 municipalités. Les quelques 1 800 000 ménages ne sont donc bien entendu pas tous interviewés. Le Tableau 2.1 montre l'évolution de la couverture de l'enquête OD depuis 1970 (Agence Métropolitaine de Transport, 2010). Pour pallier le fait qu'une petite portion de la population seulement est interrogée, des facteurs d'expansion sont utilisés pour chaque année d'enquête. Des facteurs d'expansion sont ainsi attribués aux 4.35% de ménages enquêtés en 2013. Ces facteurs ont également pour but de palier le fait que les taux de réponse ne sont pas uniformes à travers les différents segments de population. Par exemple, les ménages comprenant des personnes âgées sont souvent surreprésentés puisque ces derniers sont plus susceptibles de se trouver au domicile au moment de l'appel et de pouvoir répondre à l'interview téléphonique. Le rôle du facteur d'expansion va donc être de diminuer le poids des personnes surreprésentées et d'augmenter celui des catégories sous-représentées (MTMDET, 2017a). L'établissement de ce facteur s'appuie sur le recensement canadien, source la plus fiable pour estimer le nombre de personnes dans les différents segments de population (groupe d'âge – genre – région de domicile). La version de l'enquête OD utilisée dans ce projet est celle du MTMDET qui ajuste la population sur les estimations de l'Institut de la Statistique du Québec à l'année d'enquête à l'aide de facteurs d'expansion. L'application de ces facteurs d'expansion, aussi appelés poids d'échantillonnage, est effectuée en deux temps : premièrement un facteur d'expansion calculé suivant le nombre de ménages recensés par strate

géographique est appliqué à toutes les variables se rapportant aux ménages. Ensuite, un autre facteur calculé à partir du nombre de personnes par groupe d'âge et de genre par strate est appliqué aux variables de la personne. L'application de ces facteurs permet de diminuer le biais que peut engendrer la surreprésentation ou la sous-représentation de certains segments de population dans l'analyse des données de l'enquête OD. Cependant, celui-ci ne garantit pas une absence totale d'erreur d'échantillonnage.

Tableau 2.1 : Évolution de la couverture de l'enquête OD (Agence Métropolitaine de Transport, 2010)

	1970	1974	1978	1982	1987	1993	1998	2003	2008	2013
Population (en millier)	2484	2835	2954	2896	2930	3278	3499	3613	3940	4288
Échantillon (%) de ménages	3,78	4,78	5,31	6,98	4,68	4,65	4,64	4,70	4,10	4,35

2.5.2 Reconstitution de la population et facteurs de pondération

Les principaux facteurs qui influencent la demande de transport et qui sont pris en compte dans l'approche de prévision du MTMDET sont : la démographie, le statut, la motorisation, le profil modal, l'étalement de la période de pointe et la distribution des destinations à motif travail par secteur de domicile, genre et groupe d'âge de cinq ans (MTMDET, 2017b). Ainsi, pour « reconstituer » la population future, le facteur de pondération (de la personne) de l'enquête OD la plus récente est multiplié par le rapport : valeur projetée sur valeur courante de chaque facteur cité précédemment. Un exemple de projection du facteur de pondération pour 2036 en se basant sur l'enquête OD 2013 est donné ci-dessous :

$$F_{tsg}^{2036} = F_{tsg}^{2013} * \left(\frac{P_{tsg}^{2036}}{P_{tsg}^{2013}} \right) * \left(\frac{S_{tsg}^{2036}}{S_{tsg}^{2013}} \right) * \left(\frac{M_{tsg}^{2036}}{M_{tsg}^{2013}} \right) * \left(\frac{PM_{tsg}^{2036}}{PM_{tsg}^{2013}} \right)$$

Où P correspond à la population, S au statut, M à la motorisation, PM au profil modal, t au territoire de domicile, s au sexe et g au groupe d'âge.

À cela s'ajoute un ajustement sur l'objet déplacement pour tenir compte de l'étalement de la pointe du matin puis une redistribution des déplacements à motif travail (MTMDET, 2017b).

2.5.2.1 Démographie

La prévision de la démographie s'appuie sur le modèle ES-3. La population de chaque secteur de recensement (de Statistique Canada) est projetée suivant la mortalité, la fécondité et les migrations.

2.5.2.2 Statut

Les individus sont classés en quatre catégories : travailleur, étudiant, autre mobile et non mobile (suivant enquête OD) puis des hypothèses d'évolution futures sont posées suivant les tendances des précédentes enquêtes, toujours par groupe d'âge, genre et région du domicile.

2.5.2.3 Motorisation

Depuis peu, le facteur motorisation caractérise la personne selon la possession du permis de conduire et la motorisation de son ménage d'appartenance selon les classes de ménages suivantes : sans automobile, en « auto-négociation » (plus de personnes que de voitures) ou « auto-suffisant » (plus de voitures que de personnes). Pour la motorisation, les hypothèses sont posées par groupe d'âge, de genre, de région du domicile et de présence d'enfant.

2.5.2.4 Profil modal

Comme pour la motorisation et le statut, des hypothèses tendanciellées d'évolution de chaque profil modal parmi six (auto-conducteur seulement, transports collectifs seulement, automobile passager seulement, transport collectif multimodal, vélo seulement, autre) sont faites toujours suivant la combinaison groupe d'âge, de genre, de région du domicile, de présence d'enfant et de niveau de motorisation.

2.5.2.5 Étalement de la période de pointe

Ici, deux paramètres sont pris en compte. Premièrement le poids de chaque heure d'élargissement de la pointe par rapport à la période de pointe élargie (exemple le poids de l'heure de 5 à 6h sur une période de pointe élargie de 5 à 10h : $p = \frac{p_{5 \text{ à } 6h}}{p_{5 \text{ à } 10h}}$). Et deuxièmement l'évolution de ce poids d'une enquête à l'autre.

2.5.2.6 Distribution des déplacements à motif travail

Pour ce facteur, un modèle de Furness est appliqué pour changer les poids des déplacements et tenir compte du fait que le nombre de déplacements à motif travail va augmenter à destination des banlieues de Montréal à cause du déplacement des pôles d'emplois.

2.5.3 Exemples de prévision de la mobilité par le MTMDET

À l'aide de cette approche, le MTMDET est capable de dresser un portrait tendanciel de la mobilité future dans la Grande Région de Montréal. Par exemple, il prévoit une augmentation de la mobilité de 18,5% d'ici 2036 (voir Figure 2.1).

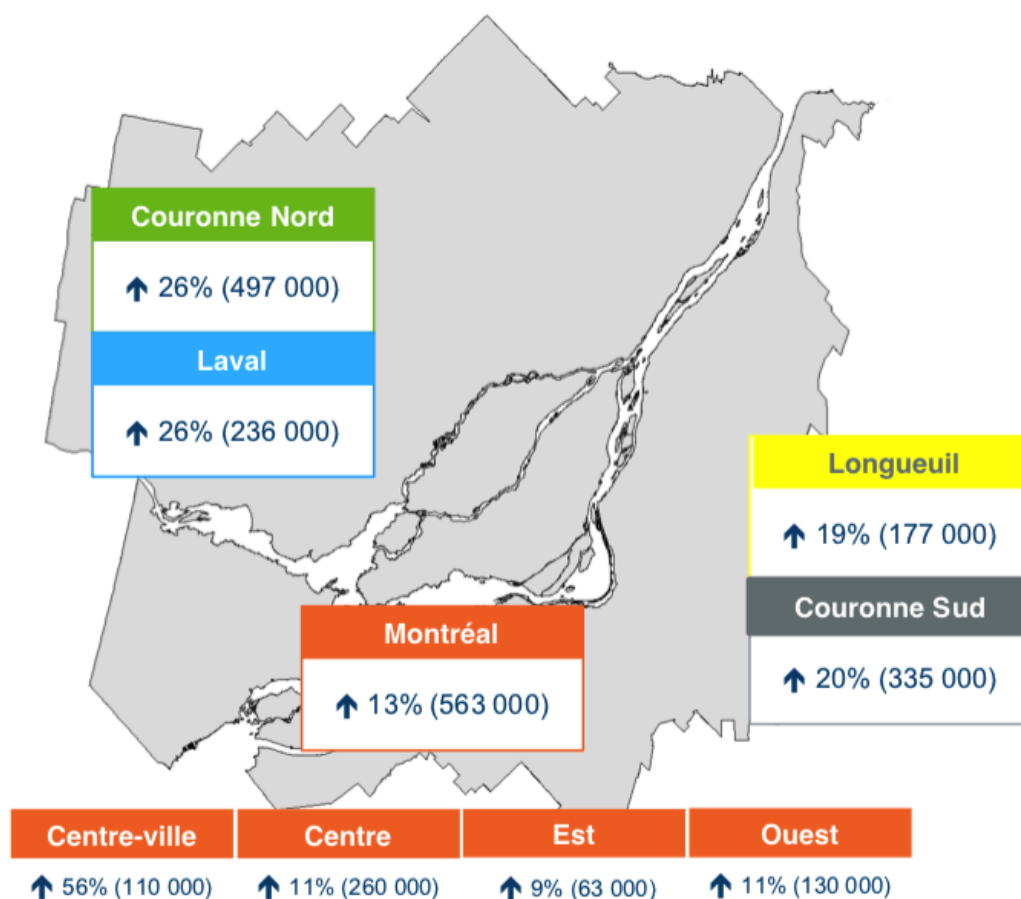


Figure 2.1 : Prévision des déplacements à l'horizon 2036 dans la GRM (MTMDET, 2017b)

De même, la part modale des déplacements auto-conducteur devrait diminuer au profit des déplacements en transports actifs et collectifs (voir Figure 2.2).

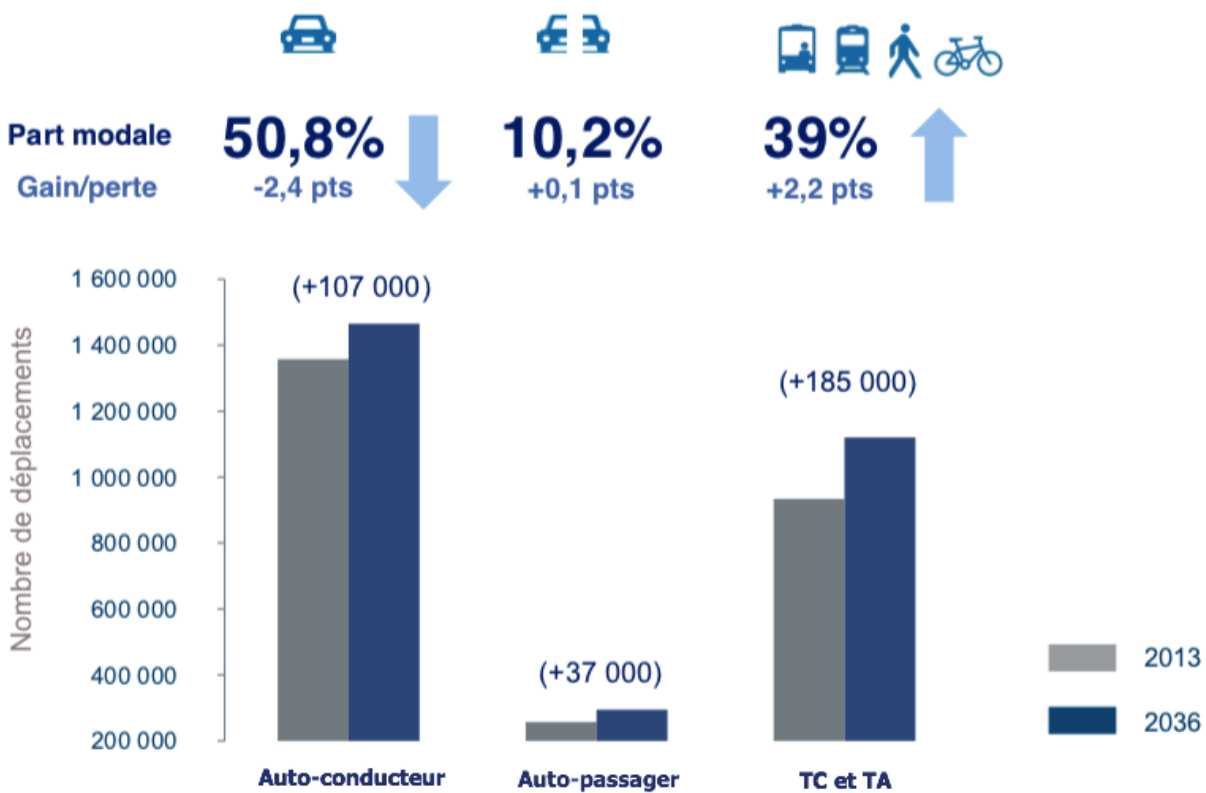


Figure 2.2 : Prévisions des parts modales à l'horizon 2036 dans la GRM (MTMDET, 2017b)

CHAPITRE 3 MÉTHODOLOGIE

Pour donner suite à l'introduction de ce mémoire, qui a posé les fondements contextuels du projet et à la revue de littérature, l'objet de ce chapitre est de présenter les grandes lignes des données et des méthodes utilisées dans cette étude. Celles-ci seront ensuite appliquées dans les chapitres 4, 5 et 6 afin de répondre au but général du projet et aux sous-objectifs définis dans la section 1.3.

3.1 Méthodologie générale

Les grandes étapes de ce projet sont présentées dans la Figure 3.1. La première consiste à sélectionner les variables de la personne et du ménage les plus pertinentes, parmi celles fournies ou calculées à partir de l'enquête OD, pour décrire les comportements de mobilité de la personne. Les comportements de déplacement sont représentés par une liste d'indicateurs. Des arbres de décision ont été utilisés. Un arbre est réalisé pour chaque indicateur et permet de déterminer les variables les plus influentes. Une fois que la sélection des variables a été faite, le but est de créer une typologie de ménages pertinente, c'est-à-dire dont les comportements de déplacement diffèrent d'un type à un autre. Le partitionnement de données (algorithme des k-moyennes) est utilisé dans cette étape. Les comportements de mobilité de chaque type de ménage sont étudiés et des tests statistiques sont appliqués pour vérifier la significativité des différences d'un groupe à l'autre. En parallèle, un modèle probabiliste d'affectation des personnes à un type de ménage est créé. Celui-ci sera ensuite utilisé pour affecter les personnes de l'enquête Origine-Destination à un type de ménage, en utilisant les variables prévisibles par les hypothèses tendanciennes formulées à partir des données de recensement seulement. Finalement, les différences de comportements de mobilité d'un type de ménage à un autre et l'affectation des personnes à ces derniers seront confrontées pour évaluer la force du modèle. Il est important de préciser que tout au long de l'étude, un point d'honneur est mis sur la transférabilité temporelle des analyses, étant donné que le but ultime de ce projet est de pouvoir utiliser les typologies pour prévoir la demande de transport future.

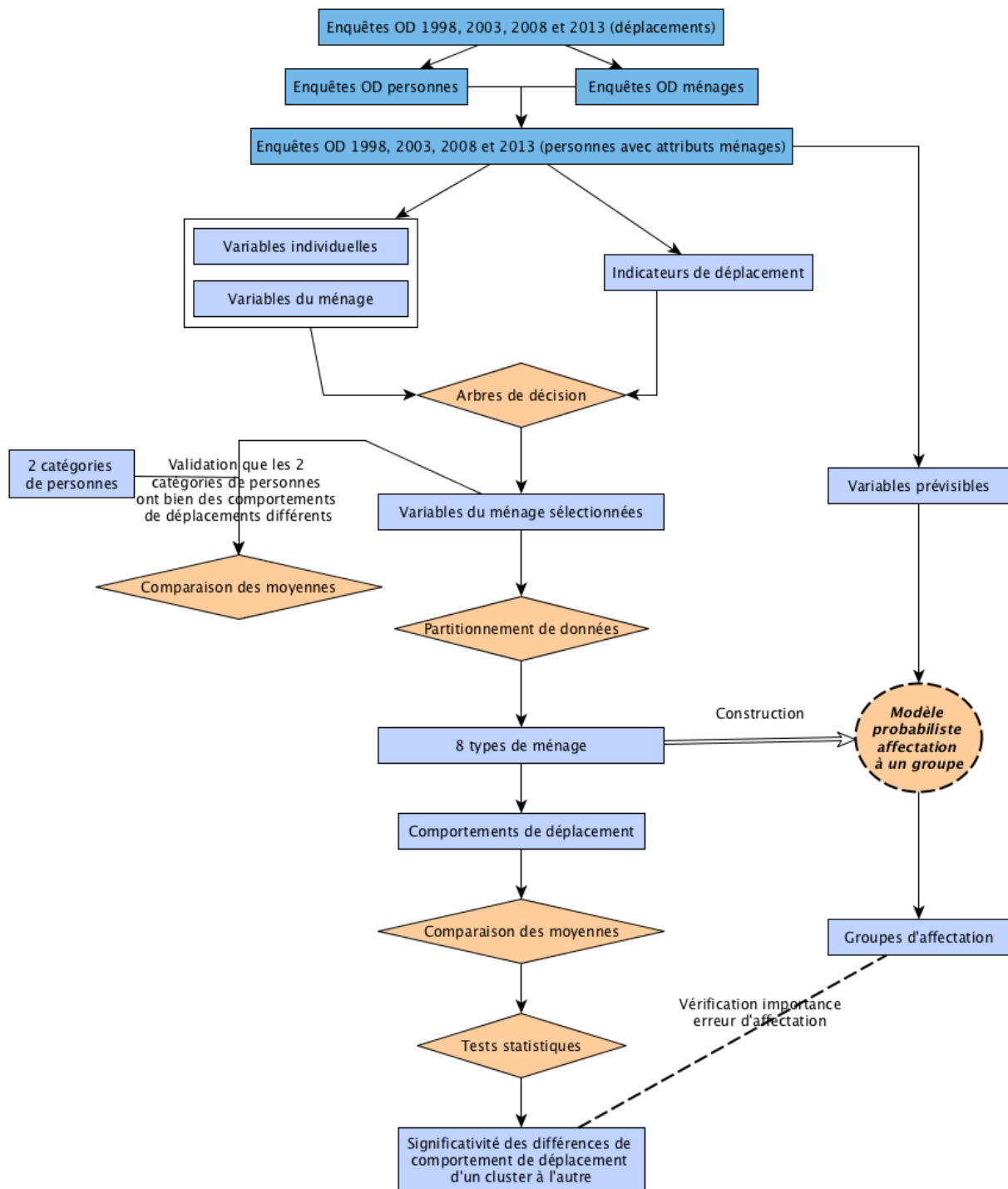


Figure 3.1 : Schéma méthodologique global

3.2 Source de données

3.2.1 Description de l'enquête OD

Les données utilisées dans cette étude proviennent des enquêtes Origine-Destination de la Grande Région de Montréal de 1998, 2003, 2008 et 2013. L'enquête OD est conduite tous les cinq ans depuis 1970 auprès des ménages de la GRM afin de recueillir leurs habitudes de déplacements. C'est une enquête téléphonique durant laquelle une personne du ménage est chargée de répondre à des questions sur sa situation personnelle (âge, genre, emploi...), celle de son ménage (nombre de personnes, nombre de véhicules...) ainsi que sur ses déplacements du jour ouvrable précédent (origines, destinations, heures de départ, modes, motifs...). La personne enquêtée doit aussi donner ces informations pour les autres membres de son foyer. L'enquête OD est conduite par un consortium d'autorités de transport de la région Montréalaise; en 2013, les partenaires de l'enquête OD étaient : Ministère des Transports de la Mobilité Durable et de l'Électrification des Transports (MTMDET), Agence Métropolitaine de transport (AMT), Société de transport de Montréal (STM), Communauté Métropolitaine de Montréal (CMM), Réseau de Transports de Longueuil (RTL), Société de transport de Laval (STL) et l'Association Québécoise du Transport (AQTr). Cette enquête fournit les détails de l'ensemble des déplacements d'un « jour moyen de semaine d'automne » d'environ 5 % de la population de la GRM.

3.2.1.1 Variables utilisées dans la présente étude

Les informations recueillies dans les enquêtes OD peuvent être réparties en trois sous-ensembles : le ménage, la personne et le déplacement. La Figure 3.2 présente les liens entre ces sous-ensembles.

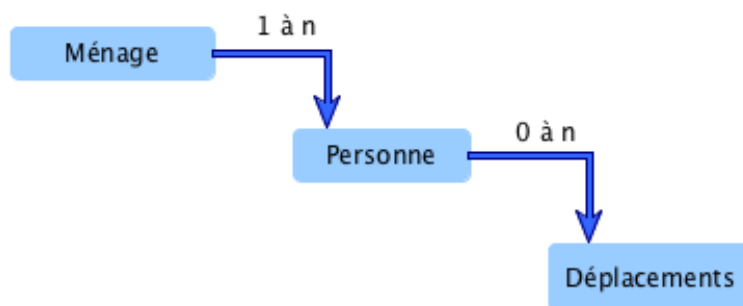


Figure 3.2 : Sous-ensembles de l'enquête OD

Bien sûr, la totalité des variables de l'enquête OD n'est pas utilisée dans cette étude; les variables retenues sont indiquées dans le Tableau 3.1.

Tableau 3.1 : Description des champs de données des enquêtes OD pertinents pour l'étude

IPERE	Numéro séquentiel unique	TLOG	Numéros de tête associés au logis
NOLOG	Numéros du logis	FACMEN/FACLOG	Facteur d'expansion ménage
SMLOG	Secteur d'analyse du logis	REGLOG	Numéro de la région du logis
XMTMLOG	Coordonnée X du logis provenant du système MTM NAD83	YMTMLOG	Coordonnée Y du logis provenant du système MTM NAD83
NBPER	Nombre de personnes dans le logis	NBVEH	Nombre de véhicules dans le logis
TPER	Numéros de tête associés à la personne	FACPER	Facteur d'expansion de la personne
NOPER	Numéro de la personne	SEXE	Sexe de la personne
AGE	Âge de la personne	GRPAGE	Groupes d'âge de 5 ans
PERCOND	Possession d'un permis de conduire	OCCPER	Occupation de la personne
MOBIL	Mobilité de la personne	NODEP	Numéros de déplacement de la personne
HREDEP	Heure de départ	GHREDEP	Groupes d'heure de départ
MODE1	Premier mode emprunté	MOTIF	Motif du déplacement

3.2.1.2 Fichiers chaînes

En plus des données fournies directement par l'enquête OD, il est possible d'obtenir des informations sur les chaînes de déplacements effectuées par les personnes de l'enquête OD. Pour cela, une table sur les chaînes de déplacements des individus est créée à partir des données de déplacement. Elle permet notamment d'obtenir le nombre de déplacements par chaîne, la durée de celle-ci, la complexité, la distance, les modes utilisés et autres attributs des chaînes... (Valiquette, 2010). Les tables de chaînes associées à chaque année (voir le Tableau 3.2 pour les champs utilisés dans les présents travaux) sont donc utilisées. À chaque personne de l'enquête OD sont associés le nombre de chaînes effectuées, le nombre de déplacements par chaîne, le

nombre de chaînes simples et le nombre de chaînes ouvertes. Cette étape est détaillée dans l'annexe E.

Tableau 3.2 : Description des champs de la base de données de chaînes des enquêtes OD pertinents pour l'étude

FEUILLET	Identification unique du ménage (équivalent NOLOG)	PERS_RANG	Rang de la personne au sein du ménage
CH_IPERE	Séquence des déplacements de la chaîne	CH_NO	Numéro de la chaîne par personne
NB_DEPL	Nombre de déplacements au sein de la chaîne	CH_NB_ACT	Nombre d'activités au sein de la chaîne

3.2.2 Faits saillants des enquêtes OD

L'analyse des quatre dernières enquêtes OD réalisées dans la Grande Région de Montréal depuis 1998 a permis de déceler et de confirmer certaines des tendances de démographie et de modification des comportements de déplacement mentionnées dans la revue de littérature ci-dessus.

Premièrement, l'augmentation de la population : à territoires comparables, l'augmentation de la population entre 1998 et 2013 approche les 15%. La population a vu son taux de croissance à son maximum entre 2003 et 2008 avec 6%. Celui-ci, même s'il a un peu baissé à partir de 2008, reste supérieur à 5% en 2013. Pour ce qui est du vieillissement de la population, l'analyse des enquêtes OD de 1998 à 2013 confirme aussi les tendances mentionnées dans la littérature. Comme l'indique la Figure 3.3, on observe entre 1998 et 2013 une constante diminution du nombre de personnes dans les catégories d'âge inférieur à 45 ans tandis que les plus de 45 ans augmentent continuellement.

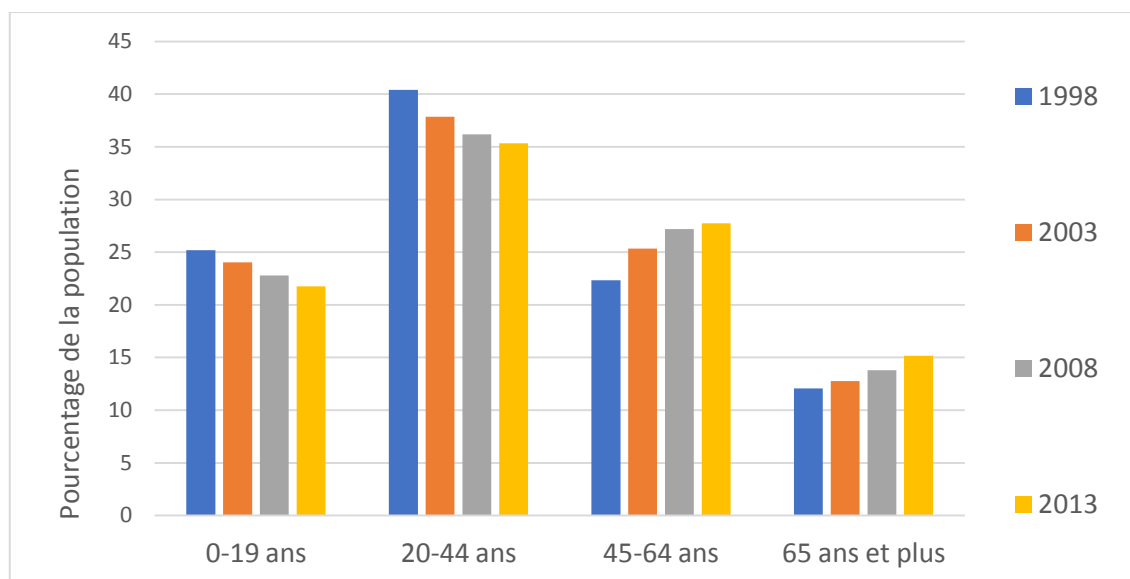


Figure 3.3 : Pourcentage de personnes par catégorie d'âge suivant l'année d'enquête

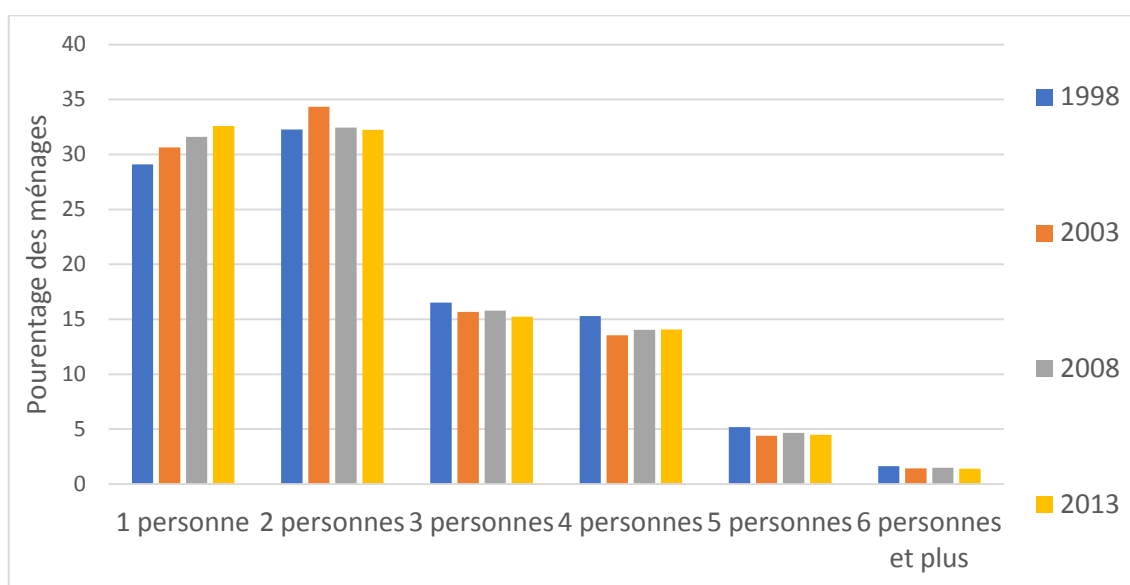


Figure 3.4 : Pourcentage de ménages par taille suivant l'année d'enquête

De même pour la taille des ménages (toujours à territoire constant); une augmentation certaine et constante des ménages d'une personne est observée entre 1998 et 2013, tandis que les ménages de trois personnes ou plus ont plutôt tendance à diminuer. Les ménages de deux personnes ont vu leur part fortement augmenter entre 1998 et 2003 puis celle-ci a diminué pour revenir, en 2013, au même taux qu'en 1998. Ceci est illustré dans la Figure 3.4.

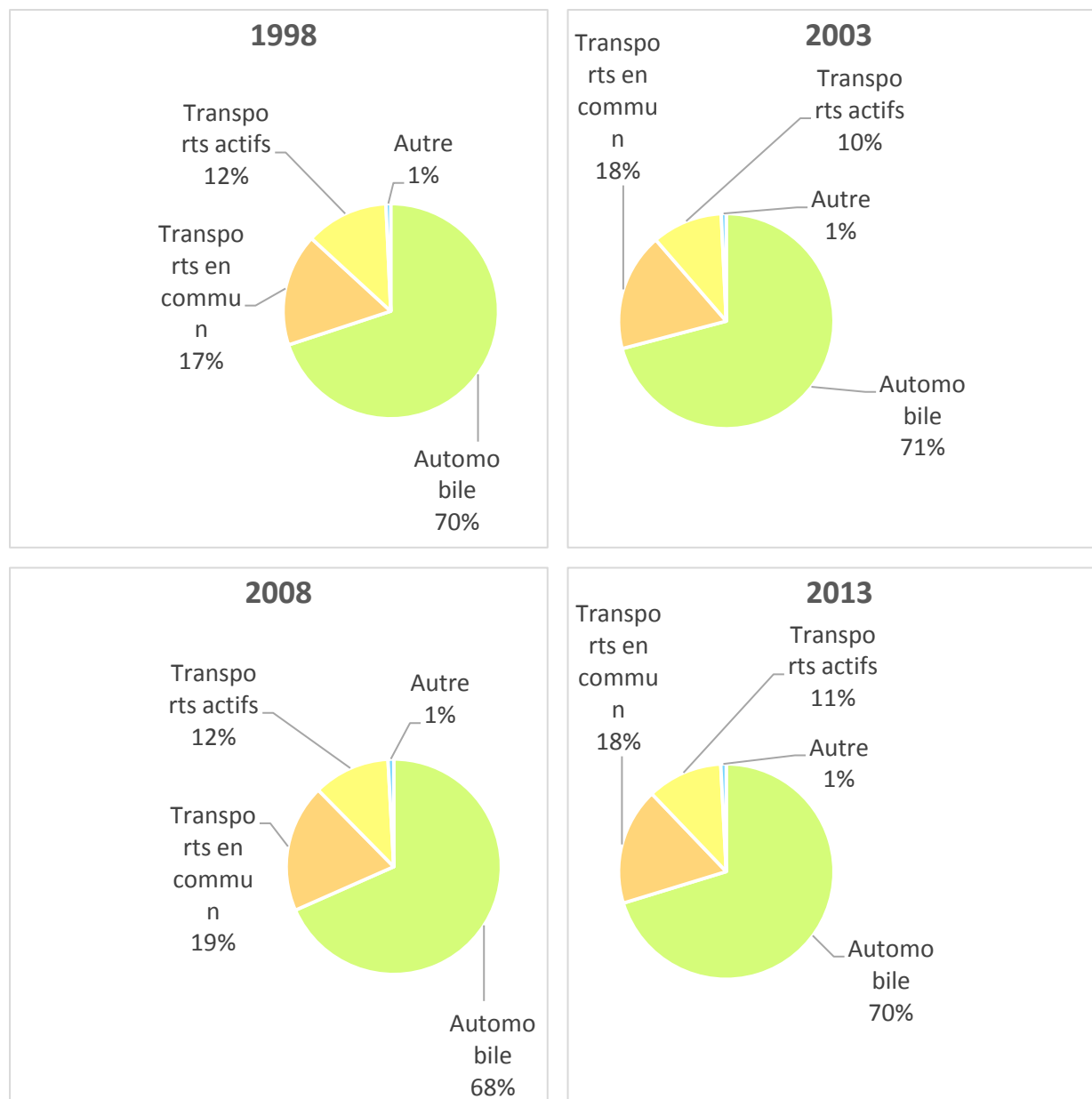


Figure 3.5 : Parts modales par année (territoire d'enquête 1998)

En ce qui concerne les déplacements des personnes, les enquêtes OD permettent effectivement de déceler des tendances, notamment l'augmentation de l'utilisation des transports en commun entre 1998 et 2008, comme l'illustre la Figure 3.5 (comparaison à territoire constant).

Enfin, il est possible d'observer quelques liens entre la structure des ménages et les déplacements des personnes grâce aux enquêtes OD. Ainsi, comme représenté dans la Figure 3.6, la présence de personnes âgées dans un ménage a une influence sur le nombre de déplacements effectués par

une personne. Les personnes habitant dans des ménages dans lesquels se trouvent des personnes âgées ont plus tendance à faire peu de déplacements, voire aucun.

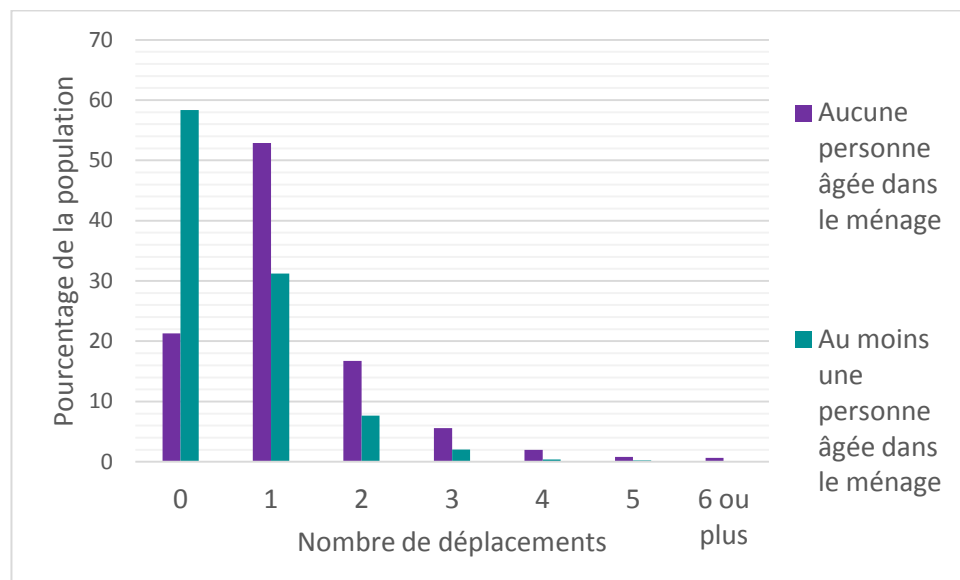


Figure 3.6 : Nombre de déplacements par personne en fonction de la présence de personnes âgées dans le ménage (enquête 2003)

À l'inverse, la présence d'enfants s'accompagne généralement d'un plus grand nombre de déplacements par personne.

3.2.3 Travail effectué sur les données

Même si les jeux de données reçues pour le projet (Enquêtes OD de Montréal : version *mtl13pv2e* pour 2013, *mtl08pv2ar_mtg* pour 2008, *mtl03p* pour 2003 et *mtl98p* pour 1998) avaient déjà subi un traitement par le MTMDET, une analyse de ces dernières était nécessaire.

3.2.3.1 Vérification des données, agrégation et comparaison entre les années d'enquête

Premièrement, les variables intéressantes pour le projet et disponibles pour les quatre années d'enquêtes ont été sélectionnées. Pour cela, un logiciel de gestion de base de données : DB Browser for SQLite a été utilisé. Bien sûr, les données ont d'abord été validées : l'absence de doublon et d'enregistrement incomplet (c'est-à-dire pour lesquels un ou plusieurs attributs n'a pas de valeur alors que cela serait attendu) a été vérifiée. Il est à préciser qu'aucune donnée n'a dû être supprimée à la suite de cette étape. Un travail d'uniformisation a également dû être effectué pour que toutes les variables aient le même nom d'une enquête à l'autre (par exemple le

facteur d'expansion de la personne porte un suffixe de l'année d'enquête dans la base de données, de même le nom de la variable heure de départ n'est plus le même en 2013). De plus, la variable REGLOG (numéro de la région du logis, de 1 à 8) n'existe pas dans l'enquête 1998. Seule la variable SMLOG (secteur d'analyse du logis, de 1 à 100) est disponible. Une correspondance entre le secteur d'analyse du logis et la région de ce dernier a donc été effectuée. Aussi, la couverture spatiale de l'enquête 2013 étant plus large que celles des années précédentes (certaines municipalités se sont ajoutées, comme par exemple Saint-Jacques-le-mineur ou bien Saint-Jean-Baptiste), lorsqu'une comparaison numérique entre les années était faite, seules les données du même territoire de l'enquête 1998 ont été conservées. Enfin, certaines variables ont été agrégées pour en faciliter l'étude. En effet, de nombreuses variables sont transformées en booléens dans la suite des travaux (par exemple pour l'occupation de la personne, quatre variables sont créées : travailleurs à temps plein (OUI ou NON), travailleurs à temps partiel (OUI ou NON), étudiant (OUI ou NON) et retraité (OUI ou NON)). L'agrégation des variables facilite cette étape. Aussi, lorsque les variables sont utilisées dans les arbres de décision, un trop grand nombre d'attributs possibles rend l'algorithme plus lourd (à réaliser et à interpréter). Le Tableau 3.4 résume, pour les variables qui ont été agrégées, les attributs avant et après agrégation. Certaines variables ayant été agrégées au début du projet n'ont finalement pas été utilisées dans les arbres ou autre étapes, mais sont tout de même répertoriées dans le tableau suivant. Pour le mode utilisé, seul le premier mode a été retenu afin de faciliter l'analyse. Il est considéré que le nombre de déplacements bi-modaux dont le premier mode n'est pas le mode majoritaire dans le déplacement n'est pas conséquent. Un vecteur de déplacement aurait pu être construit à la place.

Tableau 3.3 : Description des variables avant et après agrégation

Variable	Avant agrégation	Après agrégation
PERCOND	Oui, Non, Ne sait pas, Refus de répondre, Moins de 16 ans	Oui, Autre

Tableau 3.3 : Description des variables avant et après agrégation (suite)

OCCPER	Travailleur à temps complet, Travailleur à temps partiel, Étudiant/Élève, Retraité, Autre, 4 ans et moins, À la maison, Refus de répondre	Travailleur à temps complet, Travailleur à temps partiel, Étudiant/Élève, Retraité, Autre
PASSETC	Titre mensuel métropolitain TRAM, Titre mensuel TRAIN, Titre mensuel STM, Titre mensuel RTL, Titre mensuel STL, Titre mensuel CIT-OMIT-MRC, Enfant de 5 ans et moins, Aucun titre mensuel, Ne sait pas/Refus de répondre	Titre mensuel, Aucun titre mensuel (incluant enfant de 5 ans et moins, aucun titre mensuel et ne sait pas/Refus de répondre)
MOTIF	Travail, Rdv d'affaires, Sur la route, Étude/École, Magasinage, Loisir, Visite (amis, parenté), Santé, Reconduire quelqu'un, Chercher quelqu'un, Retour au domicile, Autre, Indéterminé/Refus de répondre/Ne sait pas	Travail (incluant travail et rdv d'affaire), Études (incluant étude et école), Magasinage, Loisir, Retour au domicile, Accompagnement (incluant reconduire quelqu'un et chercher quelqu'un)
MODE	Auto-conducteur, Auto-passager, Autobus STM, Métro, Autobus RTL, Autobus STL, Autobus CIT-OMIT-MRC, Train, Autobus scolaire, Autre bus, Taxi, Moto, Vélo, À pied, Transport adapté, Avion/train/Autobus interurbain, Point de jonction, Indéterminé	Auto, Transports en commun, Transports actifs, Autre (Taxi, moto, transport adapté, Avion/train/Autobus interurbain, Point de jonction, Indéterminé)

3.2.3.2 Création d'une base de données « personnes » intégrant les attributs du ménage

À la suite de cette phase de prétraitement des données, une base de données, dont la clé unique est l'ID du déplacement, est obtenue. Elle est prête à être utilisée pour créer une seconde base de données dont la clé unique est l'ID de la personne, incluant les caractéristiques de son ménage. À partir de là, tout le traitement de la base de données est fait à l'aide du logiciel R. La première étape pour former cette nouvelle table consiste à créer un attribut « ID personne » et un autre « ID ménage » pour chaque ligne de la base de données initiale. Le code R de cette étape est détaillé dans [l'annexe A](#). Aussi, avant de créer la base de données « personnes », certains attributs sont transformés en booléens afin de pouvoir sommer ces attributs sur le ménage (pour les attributs de la personne) ou sur la personne (pour les attributs de déplacements). C'est le cas notamment du genre, de l'occupation, du motif, du mode... De plus, de nouveaux attributs sont créés tels que le fait d'utiliser un mode doux ou non.

3.2.3.2.1 Création d'une base de données « personnes »

Une fois l'ID personne obtenu pour toutes les observations de la base de données initiale, les caractéristiques de déplacement sont agrégées pour chaque personne. C'est notamment le cas du motif (pour une personne, le nombre de déplacements pour chaque motif est compté), du nombre de déplacements, de l'heure de départ, du mode et de la distance parcourue. Avec ces nouvelles variables affectées à la personne, des indicateurs sont calculés :

- Le nombre de déplacements;
- Le nombre de déplacements en enlevant les retours au domicile;
- La distance moyenne parcourue (à vol d'oiseau);
- La part de déplacements en transports actifs (marche à pied ou vélo);
- La part de déplacements en transports en commun;
- La part de déplacements en automobile;
- Le nombre de déplacements en mode doux (transports actifs ou collectifs) sur le nombre de déplacements en mode non doux;
- Le nombre de déplacements en transports actifs sur le nombre de déplacements en auto;
- Le nombre de déplacements en transports en commun sur le nombre de déplacements en auto;
- La part de déplacements en pointe AM;

- La part de déplacements en pointe PM;
- Le nombre de chaînes complexes effectuées (une chaîne complexe est une chaîne comportant deux activités ou plus).

3.2.3.2.2 *Création d'une base de données « ménages »*

De la même manière, une autre base de données, dont l'identifiant unique est l'ID ménage, est créée. Au préalable, de nouvelles catégories d'âge sont créées pour la personne (variables booléennes : inférieur à 6 ans, entre 6 et 11 ans, entre 12 et 16 ans, entre 76 et 84 ans, supérieur à 84 ans ou autre). Pour créer la nouvelle base de données, les caractéristiques de la personne et des déplacements sont agrégées pour chaque ménage. C'est notamment le cas du genre, de l'occupation, de la possession du permis de conduire, des catégories d'âge citées précédemment et des attributs de déplacements déjà agrégés pour la personne. Les mêmes indicateurs qui ont été créés pour la personne sont ensuite créés pour le ménage.

Une fois ces deux dernières bases de données obtenues, elles sont regroupées pour former une base de données dont la clé unique est l'ID de la personne regroupant ses caractéristiques individuelles, les caractéristiques de ses déplacements et celles de son ménage. Ensuite, de petites modifications sont effectuées : les déplacements « retour » sont décomptés, le nombre de déplacements (sans retour) est agrégé (de 0 à 6 ou plus), les enfants de moins de 5 ans sont éliminés de la base de données (mais ils sont comptés dans les attributs du ménage). Enfin un travail sur les âges dans le ménage est fait pour obtenir l'âge moyen du ménage ainsi que l'amplitude maximale des âges. Finalement, voici la liste des attributs de chaque ligne de cette base de données (Tableau 3.5) :

Tableau 3.5 : Liste des attributs de la table finale

Objet	Personne	Ménage
Caractéristiques	IDPERS (int) FACPER (num) PASSTC (bool) MOBIL (bool) SEXE (cat) AGE (int) GRPAGE (cat) INF_6ANS (bool) INF_12ANS (bool) ENTRE6_11ANS (bool) ENTRE12_16ANS (bool) ENTRE76_84ANS (bool) SUP_84ANS (bool) OCCPER (cat) PERCOND (bool) MEN (bool) WOMEN (bool) MENPERCOND (bool) WOMENPERCOND (bool) FTWORK (bool) PTWORK (bool) STUDENT (bool) RETIREE (bool) MENFTWORK (bool) MENPTWORK (bool) MENSTUDENT (bool) MENRETIREE (bool) WOMENFTWORK (bool) WOMENPTWORK (bool) WOMENSTUDENT (bool) WOMENRETIREE (bool) NBDEP (int) NBDEPsanretour (cat de 0 à 6 et plus)	IDLOG (int) FACLOG (num) NBPER (nt) NBVEH (int) XMTMLOG (num) YMTMLOG (num) REGLOG (cat) SMLOG (cat) INF_6ANS (int) INF_12ANS (int) ENTRE6_11ANS (int) ENTRE12_16ANS (int) ENTRE76_84ANS (int) SUP_84ANS (int) AGEMEAN (num) AGESD (num) AGEMIN (int) AGEMAX (int) AGEAMPMAX (int) MEN (int) WOMEN (int) MENPERCOND (int) WOMENPERCOND (int) FTWORK (int) PTWORK (int) STUDENT (int) RETIREE (int) MENFTWORK (int) MENPTWORK (int) MENSTUDENT (int) MENRETIREE (int) WOMENFTWORK (int) WOMENPTWORK (int) WOMENSTUDENT (int) WOMENRETIREE (int) PASSTC (int) ACCESAUTO (num) NBDEP (int) NBDEPsanretour (cat de 0 à 6 et plus)
Motif des déplacements	TRAVAIL (int) ETUDE (int) MAGASINAGE (int) LOISIR (int) RETOUR (int) CHERCHER_RACCOMPAGNER (int)	

Tableau 3.5: Liste des attributs de la table finale (suite)

Heure de départ des déplacements	hre1 (int) hre2 (int) hre3 (int) hre4 (int) hre5 (int) hre6 (int) hre7 (int) PARTPOINTEAM (num) PARTPOINTEPM (num)	PARTPOINTEAM (num) PARTPOINTEPM (num)
Mode des déplacements	AUTO (int) TC (int) TA (int) AUTRE (int) DOUX (int) NONDOUX (int) PARTTA (num) PARTTC (num) PARTAUTO (num) DsurND (num) TAsurAUTO (num) TCsurAUTO (num)	AUTO (int) TC (int) TA (int) AUTRE (int) DOUX (int) NONDOUX (int) PARTTA (num) PARTTC (num) PARTAUTO (num) DsurND (num) TAsurAUTO (num) TCsurAUTO (num)
Chaines effectuées	CHAINE_SIMPLE (int) CHAINE_CONTRAINTTE (int) CHAINE_OUVERTE (int) CHAINE_COMPLEXE (int)	CHAINE_SIMPLE (int) CHAINE_CONTRAINTTE (int) CHAINE_OUVERTE (int) CHAINE_COMPLEXE (int)

L'ensemble des commandes effectuées pour la création de la base de données « personnes » et la base de données « ménages », ainsi que pour la correspondance des âges dans le ménage et des chaînes de déplacement de chaque personne sont détaillées dans [l'annexe B](#).

3.3 Indicateurs

3.3.1 Indicateurs de mobilité

Les indicateurs de mobilité permettent de décrire les comportements de déplacement des personnes pour une journée (la journée d'enquête). Le choix des indicateurs de déplacement pour la personne a été effectué principalement en fonction de ce qui a été trouvé dans la littérature. Le nombre d'indicateurs à représenter a été limité tout en gardant ceux qui paraissaient les plus pertinents pour décrire les comportements de déplacement des personnes. Les indicateurs finalement retenus sont les suivants :

- Être mobile ou non;
- Nombre de déplacements par personne (six variables binaires, une pour chaque nombre de déplacements de 1 à 6 et plus);
- Faire au moins une chaîne complexe ou pas;
- Part des déplacements auto (déplacements dont le mode dans l'enquête OD est répertorié comme auto-conducteur ou auto-passager);
- Part des déplacements en pointe matinale (déplacements effectués entre 6:00 AM et 9:00 AM).

3.3.2 Variables explicatives

Les variables explicatives sont des caractéristiques de la personne ou bien du ménage, disponibles dans la base de données créée comme précédemment décrit. Le choix des variables à inclure dans les modèles s'est fait selon plusieurs critères. Les variables faisant référence à des caractéristiques pouvant facilement être projetées (âge et sexe, grâce au modèle de projection de la population du MTMDDET) sont préférées tandis que des variables comme la possession du permis de conduire qui sont plus difficilement prévisibles sont retirées. Aussi, un choix est fait entre les variables trop corrélées (l'une d'elles est éliminée), comme c'est le cas, par exemple, entre l'âge et le statut. Ensuite, différentes catégories pour représenter l'âge des personnes dans le ménage ont été testées (intervalles d'âges de 5 ans, intervalles d'âges plus grands, intervalles d'âges basés sur un arbre représentant le nombre de déplacements en fonction de l'âge...). Il en ressort qu'avoir trop de catégories n'est pas une bonne chose, il a été préféré de chercher une ou deux variables résumant au mieux la dispersion des âges dans le ménage et d'y ajouter la présence d'enfants et de personnes âgées. Les variables finalement retenues sont les suivantes :

- Variables de la personne :
 - Genre
 - Âge
- Variables du ménage :
 - Région du domicile (huit régions différentes de l'enquête OD)
 - Accès à l'automobile (rapport du nombre de voitures sur le nombre de personnes de 16 ans et plus dans le ménage)
 - Âge moyen
 - Amplitude maximale des âges
 - Nombre de personnes
 - Nombre d'hommes
 - Nombre de femmes

- Présence d'enfants de moins de 12 ans
- Présence d'enfants entre 12 et 16 ans
- Présence de personnes entre 76 et 84 ans
- Présence de personnes de plus de 84 ans

La corrélation entre les variables n'a pas d'influence sur le pouvoir des arbres, cependant il reste intéressant de regarder quelles variables utilisées sont corrélées entre elles afin de déterminer si l'utilisation de certaines est redondante. Le Tableau 3.7 présente la matrice de corrélation des variables sélectionnées. Dans ce tableau, plus la couleur de la cellule tend vers le rouge, plus les variables sont corrélées et inversement lorsque la couleur de la cellule tend vers le vert. Cette analyse a été conduite plusieurs fois avant la sélection finale des variables dans le but d'éliminer des variables de catégories d'âge. C'est ainsi, par exemple, que la « présence d'enfants de moins de 12 ans » a été choisie plutôt que la combinaison des deux variables « présence d'enfants de moins de 6 ans » et « présence d'enfants entre 6 et 12 ans ». Parmi les caractéristiques de la personne et du ménage finalement choisies, présentées dans le Tableau 3.7, il est possible de remarquer que les variables du nombre de personnes dans le ménage, de l'âge moyen et de l'amplitude maximale des âges semblent corrélées à plusieurs autres variables dans l'analyse. Cependant, celles-ci seront conservées dans la composition des arbres de décision puisque la corrélation des variables ne réduit pas la qualité de l'arbre et que ces dernières ne représentent pas la même chose.

Tableau 3.7 : Matrice de corrélation des variables sélectionnées pour les arbres de décision

	Genre	Région domicile	Nb hommes	Nb femmes	NB personnes	Âge	Accès auto	Âge moyen	Âge amp max	Pres -12 ans	Pres 12-16 ans	Pres 76-84 ans	Pres +84 ans
Genre	1												
Région domicile	0	1											
Nb hommes	-0,24	0,04	1										
Nb femmes	0,24	0,02	-0,05	1									
NB personnes	-0,01	0,04	0,7	0,67	1								
Âge	-0,02	0	-0,35	-0,37	-0,52	1							
Accès auto	-0,01	0,35	-0,2	-0,27	-0,34	0,26	1						
Âge moyen	-0,01	0	-0,44	-0,45	-0,65	0,76	0,34	1					
Âge amp max	0,02	-0,02	0,47	0,52	0,71	-0,49	-0,31	-0,6	1				
Pres -12 ans	0	-0,01	-0,36	-0,37	-0,52	0,46	0,36	0,64	-0,48	1			
Pres 12-16 ans	-0,01	-0,03	-0,29	-0,32	-0,44	0,31	0,23	0,34	-0,36	0,1	1		
Pres 76-84 ans	0	0,03	0,1	0,07	0,12	-0,29	0,02	-0,38	0,08	-0,13	-0,08	1	
Pres +84 ans	0	0,02	0,05	0,01	0,04	-0,15	0,03	-0,22	-0,04	-0,07	-0,05	0,14	1

3.4 Arbres de décision

La section suivante a pour but d'expliquer la méthode des arbres de décision appliquée pour ce projet grâce au logiciel R et de faciliter l'interprétation des figures qui seront présentées dans le Chapitre 4.

Le but de l'utilisation des arbres de décision est d'évaluer quelles sont les variables de l'enquête OD représentant les caractéristiques de la personne et du ménage les plus déterminantes pour les comportements de déplacement individuels. C'est en quelque sorte une façon de déterminer une hiérarchie entre les variables, pour pouvoir ensuite faire une sélection de celles qui ont le plus grand pouvoir explicatif quant aux divers comportements de mobilité.

L'arbre de décision est une méthode plutôt populaire du fait de sa capacité à produire une connaissance simple et directement utilisable, à la portée de tout le monde (Rakotomalala, 2005). C'est une méthode de fouille systématique de données (*data mining*) qui permet de créer un ensemble de feuilles et de branches à partir de règles de segmentation.

Ici, la méthode des arbres de décision à partitionnements récursifs est utilisée. Les arbres sont dans cette étude réalisés avec la fonction *rpart()* pour *Recursive PARTitioning*, du logiciel R. Cette dernière utilise de nombreuses notions provenant du livre CART (*Classification And Regression Trees*) de Breiman et Friedman (1984), une référence pour les arbres de décision et de classification. Le terme récursif est utilisé, car l'algorithme fonctionne de la manière suivante (Therneau & Atkinson, 1997) :

- D'abord, l'ensemble des données est séparé suivant la variable qui est trouvée comme la « meilleure » pour séparer les données en deux groupes;
- Ce processus est ensuite appliqué séparément à tous les sous-groupes et cela de manière récursive jusqu'à ce qu'un groupe atteigne la taille minimale (spécifiée) ou bien qu'aucune amélioration ne puisse plus être apportée;
- Enfin, l'algorithme utilise une validation croisée pour réduire la taille de l'arbre.

La commande *rpart* permet de faire des arbres de décision pour des variables, ici les indicateurs de mobilité, à la fois de type continue et de type catégorique (l'algorithme effectue dans ce cas un arbre de régression). De même, les variables explicatives peuvent être continues ou bien catégoriques. Dans le cas d'un arbre de décision, pour déterminer la « meilleure » variable pour séparer les données à chaque nœud, l'indice de Gini est utilisé. Il mesure l'impureté d'un nœud. L'impureté d'un nœud *A* est définie comme ceci :

$$I(A) = \sum_{i=1}^N p_{iA} * (1 - p_{iA})$$

Avec *N* le nombre de classes auxquelles les données peuvent être assignées et p_{iA} la proportion des données au nœud *A* appartenant à la classe *i*.

Cette impureté est calculée à chaque nœud pour le nœud en question et pour les deux nœuds « enfants » qui seraient créés, ceci pour chaque variable de séparation. La variable qui donne la plus grande réduction d'impureté ΔI est choisie.

$$\Delta I = p(A)I(A) - p(A_d)I(A_d) - p(A_g)I(A_g)$$

Avec A_d et A_g les deux nœuds enfants (d pour droite et g pour gauche).

Dans le cas d'un arbre de régression, le choix de la variable qui sépare les données en deux groupes à un nœud se fait avec une analyse de variance.

$$\Delta I = SS_T - (SS_d + SS_g)$$

Avec SS la somme des carrés du nœud concerné :

$$SS = \sum_{i=1}^N (y_i - \bar{y})^2$$

D'autres paramètres peuvent être précisés dans la fonction. Par exemple, la taille minimum des feuilles, qui indique le nombre minimum d'observations qui doivent être recensées à un nœud pour que celui-ci soit divisé (plus elle est grande, moins il y aura de feuilles terminales et ainsi de niveaux hiérarchiques). Il est également possible de modifier le paramètre de complexité qui régule le processus d'élagage de l'arbre. La valeur du paramètre de complexité détermine le gain minimum qui doit être obtenu pour faire une division supplémentaire. Cela permet donc de déterminer la taille optimale de l'arbre pour laquelle l'erreur de prédiction est la plus faible.

À chaque nœud terminal, l'algorithme détermine la valeur de la variable à expliquer la plus fréquente au sein d'une feuille et affecte cette dernière à toutes les observations classées dans la feuille. C'est ainsi qu'apparaît un taux d'erreur (nombre d'observations mal classées, c'est à dire pour lesquelles la valeur prédite ne correspond pas à la valeur observée, sur le nombre d'observations totales). Ce dernier permet d'évaluer la performance de l'arbre.

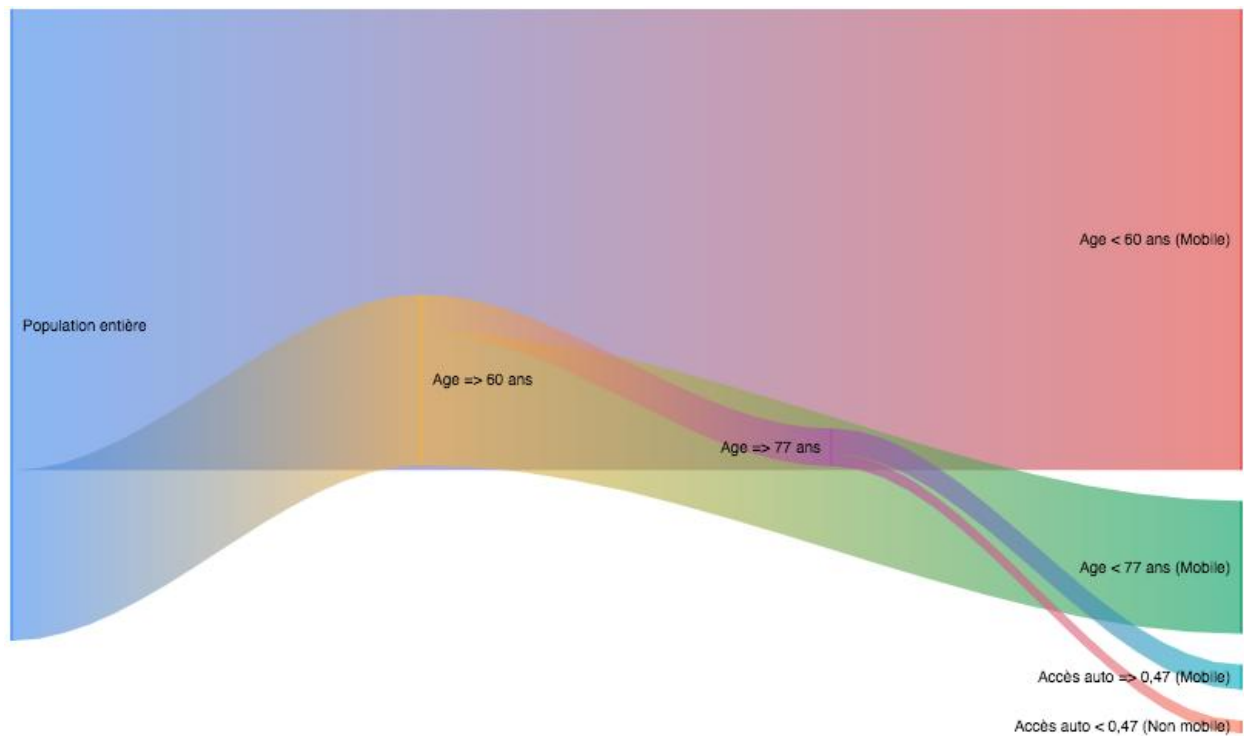


Figure 3.7 : Illustration d'un arbre de décision : être non mobile ou pas (enquête 2013)

La Figure 3.7 illustre un arbre de décision. Au sommet, le nœud principal regroupe 100% de la population. La première variable qui sépare la population est l'âge de la personne. Deux groupes sont formés : les moins de 60 ans et les personnes de 60 ans ou plus. Les personnes de 60 ans ou plus sont ensuite séparées en deux groupes par la variable âge de nouveau : les moins de 77 ans et les 77 ans ou plus. Pour ce dernier groupe, la variable de l'accès à l'automobile sépare enfin le groupe en deux. Quatre « feuilles » se distinguent donc :

- Les moins de 60 ans (qui représentent 73% de la population) et pour qui la réponse « Être mobile », est préférée.
- Les personnes de 60 ans ou plus et moins de 77 ans (21%) qui sont également prédites comme mobiles.
- Les personnes de 77 ans ou plus ayant un accès à l'auto dans leur ménage de 0,47 ou plus (4% de la population) seront mobiles aussi selon l'arbre de décision.
- Et enfin les personnes de 77 ans ou plus ayant un accès à l'auto inférieur à 0,47 (2% de la population) qui eux ont plutôt tendance à être immobiles.

Outre l'arbre de décision illustré ci-dessus, il est possible d'obtenir le taux d'erreur et la matrice de confusion. Dans l'exemple donné ci-dessus, le taux d'erreur est de 0,1715 avec 30 678 personnes mal classées sur 178 908. Ceci est illustré dans la Figure 3.8 qui présente la matrice de

confusion de l'arbre. Une nuance de bleus est utilisée pour représenter la fréquence normalisée des observations. Plus le bleu est clair plus il y a d'observations. Ainsi une matrice de confusion parfaite serait composée d'une diagonale bleu clair alors que le reste serait en bleu foncé. [L'annexe C](#) présente les étapes de la construction d'un arbre et de sa matrice de confusion.

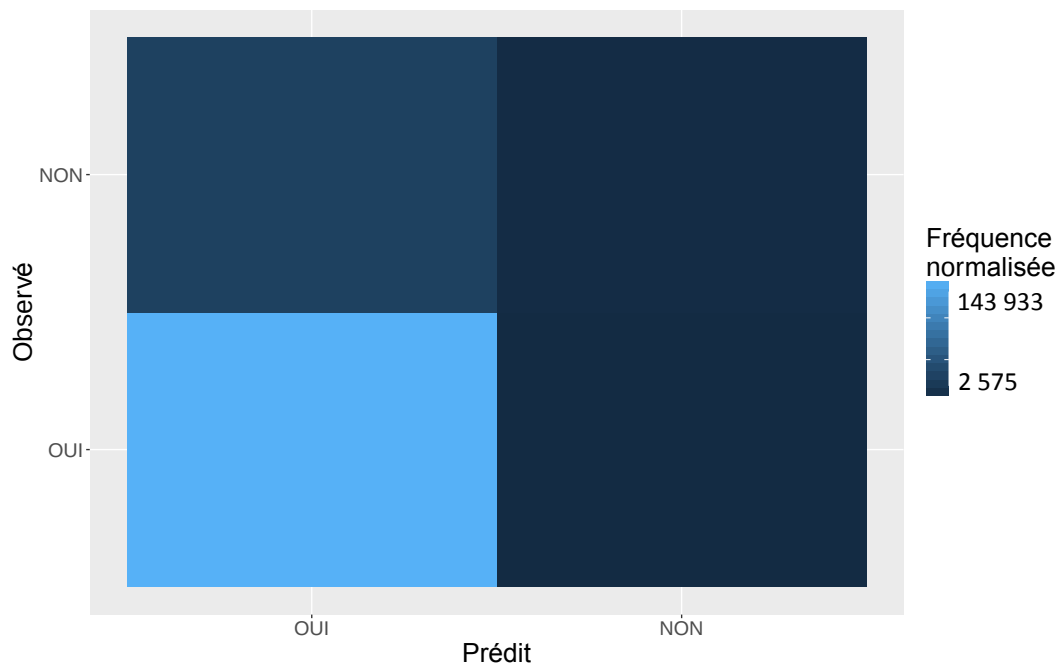


Figure 3.8 : Matrice de confusion arbre de décision : être non mobile ou pas (enquête 2013)

Enfin, l'algorithme donne l'importance des variables. C'est principalement ce paramètre qui servira à la sélection des variables dans la présente étude. En effet, la simple observation de l'arbre donné par l'algorithme ne permet pas réellement de déterminer quelles variables ont le plus d'influence pour un indicateur donné. L'importance de chaque variable est mesurée avec la diminution de la précision du modèle lorsque celle-ci en est enlevée. Les décisions de division de l'arbre sont, elles, prises en fonction de l'impureté. Or la pureté ne garantit pas la précision. En effet, une variable peut apparaître comme la deuxième meilleure pour séparer les données à tous les nœuds et donc être importante pour le modèle, sans jamais être une variable de séparation à aucun nœud. Ainsi les variables aux nœuds ne sont pas forcément les variables les plus importantes pour le modèle complet. Le Tableau 3.8 regroupe l'importance des variables du modèle présenté ci-dessus. Ces importances sont bien sûr relatives et pourraient être données sous forme de pourcentage pour une interprétation plus aisée.

Tableau 3.8 : Importance des variables de l'arbre de décision : être non mobile ou pas (enquête 2013)

Variable	Âge	Âge moyen	Pres 76-84 ans	Amp max	Accès auto	Nb personnes	Pres +84 ans	Nb hommes	Région domicile	Nb femmes	Pres 12-16 ans
Importance	5 146	4 083	1 157	802	528	508	461	86	65	35	3

3.5 Méthodes de classification

Une méthode de groupement systématique (*clustering*) est aussi utilisée dans le cadre de ce projet. Elle permet de classer les personnes en différents groupes homogènes selon les caractéristiques du ménage auquel elles appartiennent. L'algorithme utilisé est issu des méthodes de fouille systématique de données (*data mining*). Il s'agit de la méthode des k-moyennes. Cette dernière sépare l'ensemble des données en k groupes de façon à maximiser l'homogénéité à l'intérieur des groupes et l'hétérogénéité entre les groupes (Saunier, 2016). Autrement dit, minimiser la distance d'un point à la moyenne de son groupe. Ainsi, pour un ensemble de points $(x_1, x_2, \dots, x_n)$, on cherche à partitionner les n points en k ensembles $(S_1, S_2, \dots, S_k)$ tel que :

$$\min \sum_{i=1}^k \sum_{x_j \in S_i} d(x_j, \mu_i)$$

Où μ_i est le barycentre des points dans S_i .

Cet algorithme est effectué sous R et la distance euclidienne est utilisée. De ce fait, les variables catégoriques, qui ne sont pas prises en compte dans l'algorithme des k-moyennes, doivent être remplacées. Les nombres de personnes de moins de 12 ans et de 76 à 84 ans sont donc préférés (à la place de la présence de ces catégories de personnes). Les caractéristiques du ménage utilisées pour la formation des groupes sont donc toutes de type numérique (entier ou non), certaines peuvent aller jusqu'à 99 (l'âge moyen par exemple), tandis que pour le nombre de personnes, d'enfants ou de personnes âgées, il ne dépasse pas 16. Les ordres de grandeur varient donc beaucoup d'une variable à l'autre, or, la distance euclidienne est sensible aux variations qui sont du même ordre de grandeur. Si ce n'est pas le cas pour les variables utilisées, une variable peut « dominer » une autre dont l'effet sera moins pris en compte dans la création des groupes. Le fait de donner plus d'importance à certaines variables peut être intéressant. Ainsi, plusieurs

essais ont été effectués pour le regroupement des données. Le premier est effectué sur les données brutes, sans normalisation, et donne donc plus d'importance à l'âge moyen du ménage et à l'amplitude maximale des âges. D'autres essais ont été effectués en normalisant les données, afin qu'elles aient toutes la même importance. Premièrement, le « z-score » est utilisé. Il permet de transformer les variables afin qu'elles aient toutes une moyenne de 0 et une variance de 1 (Mohamad & Usman, 2013). La formule du « z-score » est définie comme suit :

$$Z(X) = \frac{X - \bar{X}}{\sigma}$$

Avec $\bar{X}$ la moyenne et σ l'écart type de la variable concernée.

Deuxièmement, la normalisation « min-max » a été appliquée. Elle permet d'obtenir des valeurs dans l'intervalle [0;1] (Mohamad & Usman, 2013). Ainsi, pour les variables du nombre de personnes dans le ménage, de l'âge moyen et de l'amplitude maximale des âges, la formule appliquée est la suivante :

$$MM(X) = \frac{X - X_{min}}{X_{max} - X_{min}}$$

Pour les autres, les variables catégoriques de présence d'enfants ou de personnes âgées sont préférées; elles prennent la valeur 0 ou 1.

Ainsi, avec les deux normalisations, toutes les variables utilisées ont le même ordre de grandeur et auront le même poids dans l'application de l'algorithme des k-moyennes.

Pour l'algorithme des k-moyennes, il est nécessaire de spécifier le nombre de groupes à créer. Ce dernier a été choisi par « essais – erreur ». Huit types de ménages seront donc créés.

L'algorithme des k-moyennes appliqué avec R permet à l'utilisateur d'obtenir le nombre de groupes souhaité, la taille, ainsi que les moyennes de chacun d'entre eux et enfin, un vecteur qui associe à chaque donnée intégrée à l'opération, le groupe auquel elle appartient. Finalement, le rapport de la somme des carrés entre les groupes sur la somme des carrés totale (entre les groupes et à l'intérieur des groupes) permet d'évaluer la classification. Plus ce rapport est grand, plus la variance du jeu de données est expliquée par des différences entre les groupes comparativement à celle intra-groupes et donc meilleure est la classification.

Le code utilisé pour créer une typologie (sans normalisation) et comparer les comportements de déplacement d'un groupe à un autre est donnée en [annexe D](#).

3.6 Tests statistiques

Une fois les types de ménages constitués, il est nécessaire de vérifier leur pertinence en validant que les comportements de déplacement des personnes diffèrent d'un type de ménage à un autre. Les moyennes et distributions des comportements de déplacement seront donc confrontées d'un échantillon à un autre grâce à divers tests statistiques.

3.6.1 Comparaison des moyennes

Pour les moyennes, le test de Tukey, également appelé HSD (pour Honestly Significant Difference) sera appliqué. Ce dernier a pour but de distinguer, parmi plusieurs échantillons, lesquels se distinguent significativement des autres. Le test calcule, pour chaque paire d'échantillons, la statistique suivante :

$$Q_{ij} = \frac{|M_{ki} - M_{kj}|}{\sqrt{\frac{M_{intra}}{n}}}$$

Où k est le nombre d'échantillons distincts, n le nombre d'observations dans les échantillons, i et j , les deux échantillons à confronter, et M_{intra} la moyenne des inerties intra-groupe (Desgraupes, 2012). Dans le cas où tous les échantillons n'ont pas la même taille, n est la moyenne harmonique des effectifs dans chaque échantillon. Cette statistique est ensuite comparée à des valeurs critiques de Q (dépendant du nombre d'échantillons et du degré de liberté). Si la valeur de Q calculée est plus grande que Q_{cr} l'hypothèse H_0 (toutes les moyennes sont égales) ne peut pas être validée. Dans cette étude, le test de Tukey est appliqué grâce aux fonctions `aov()` (application de l'ANOVA) et `TukeyHSD()` de R.

3.6.2 Comparaison des distributions

Enfin, pour la comparaison des distributions, du fait du grand nombre d'observations dans chaque cluster, aucun test ne permet la comparaison multiple. Le test de Anderson-Darling est habituellement utilisé pour cela, mais il est moins performant pour des échantillons de 1000 observations et plus. Le test de Kolmogorov-Smirnov est donc choisi pour comparer les

distributions des clusters deux à deux. Ce test est basé sur les fonctions de répartition (Desgraupes, 2014). Ainsi, l'hypothèse H_0 stipule que les distributions empiriques des deux échantillons comparés sont identiques. La méthode est donc de calculer l'écart maximum qui existe entre les deux fonctions de répartitions empiriques, soit :

$$D = \max |F_i(x) - F_j(x)|$$

Avec F_i et F_j les fonctions de répartitions des échantillons i et j respectivement. La Figure 3.9 illustre un exemple simple de l'écart maximum D .

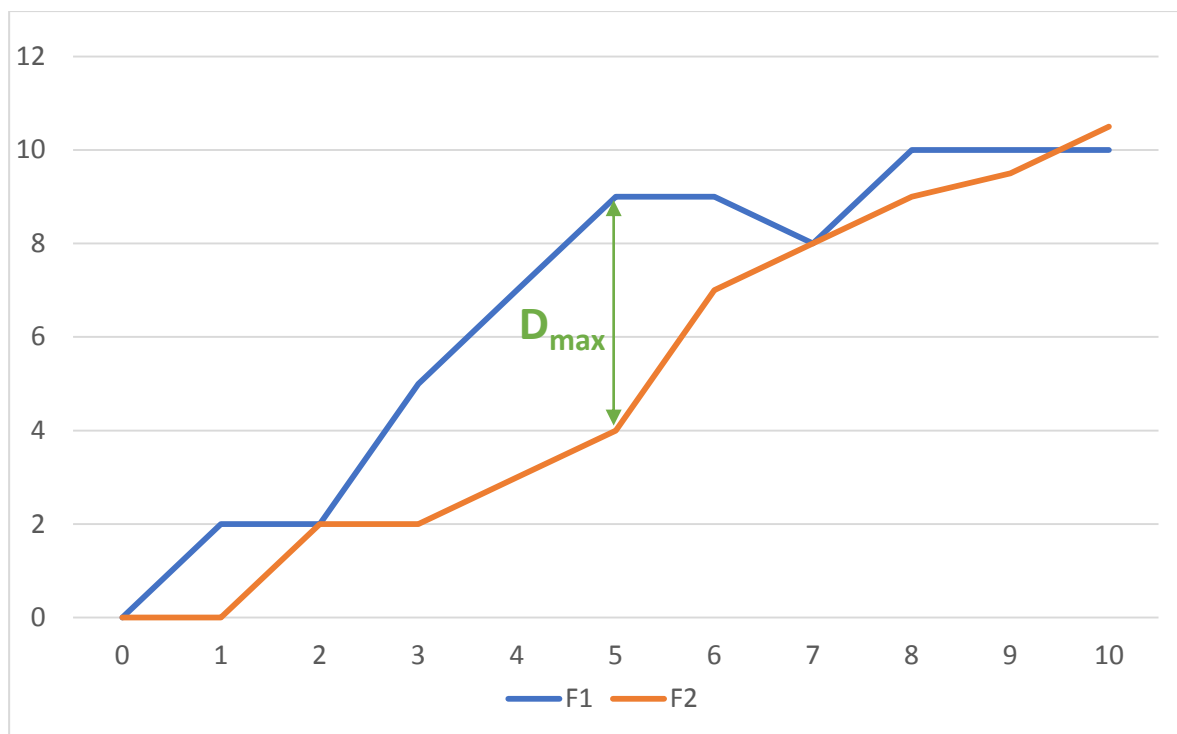


Figure 3.9 : Exemple d'écart maximal entre deux fonctions de répartition

Similairement au test de Tukey, la statistique D obtenue est comparée à une valeur critique donnée par des tables suivant l'effectif des échantillons et le seuil de risque accepté. Si D est supérieure à D_{cr} alors l'hypothèse H_0 ne peut pas être validée. La fonction `ks.test()` de R permet d'appliquer le test de Kolmogorov-Smirnov à deux groupes de données dont les distributions doivent être comparées.

3.7 Modèle logit multinomial

Pour le modèle d'affectation des personnes à un type de ménage, le modèle logit multinomial sera choisi, étant donné que dans ce dernier, la variable expliquée peut prendre plus de deux modalités (ici les modalités correspondent aux types de ménages auxquels les personnes appartiennent, il y a donc huit possibilités). Dans ce modèle, la probabilité que l'individu i choisisse la modalité j peut être écrite de la façon suivante :

$$Prob(y_i = j) = \frac{\exp(x_i \beta_j)}{1 + \sum_{k=1}^m (x_i \beta_k)}$$

Dans cette équation, les β_j peuvent varier selon les modalités de j (Hurlin, 2003). Il est à noter que les paramètres s'interprètent comme les écarts au référentiel, c'est-à-dire au paramètre de modalité 0. Dans le modèle logit multinomial, la propriété IIA doit être respectée : « le rapport de deux probabilités associées à des événements particuliers est indépendant des autres événements », en d'autres mots, la préférence pour une alternative doit être indépendante de la préférence pour une autre alternative.

Ensuite, pour parvenir à estimer les paramètres du logit multinomial, il faut maximiser la log-vraisemblance associée au modèle, c'est-à-dire, maximiser la log-vraisemblance par rapport aux vecteurs de paramètres $(\beta_1, \beta_2, \dots, \beta_m)$. Ainsi, si on considère la modalité 0 comme étant la référence, on a :

$$\ln\left(\frac{P(y = j)}{P(y = 0)}\right) = a_j + b_j * x_1 + c_j * x_2 \dots$$

Et :

$$\ln\left(\frac{P(y = k)}{P(y = 0)}\right) = a_k + b_k * x_1 + c_k * x_2 \dots$$

et ainsi de suite.

Avec $x_1, x_2, \dots$ les variables explicatives du modèle.

Ainsi, le logarithme du rapport de la probabilité d'appartenir à un groupe j plutôt qu'au groupe de référence augmentera de b_j pour une augmentation d'une unité de x_1 . Le rapport des probabilités de choisir une modalité plutôt qu'une autre est souvent référée comme le « risque

relatif ». En prenant l'exposant de l'équation, il est donc possible d'obtenir ce risque relatif. La fonction *multinom()* du package *nnet* de R est utilisée pour effectuer le modèle logit multinomial. Les coefficients affectés à chaque attribut pour chaque modalité en sont tirés, mais ces derniers sont difficiles à interpréter. Pour favoriser l'interprétation de la performance du modèle, la matrice de confusion est obtenue. La matrice de confusion représente la modalité observée versus la modalité prédite. Le code utilisé pour créer le modèle est donné en [annexe E](#).

CHAPITRE 4 SÉLECTION DES VARIABLES À INTÉGRER À LA TYPOLOGIE

4.1 Variables représentant les âges dans un ménage

De nombreuses possibilités ont été étudiées pour évaluer les variables représentant le mieux les différents âges observés au sein d'un ménage. Ainsi, la présence et le nombre de personnes appartenant à un intervalle d'âges ont été considérés : présence de personnes entre 0 et 4 ans, présence de personnes entre 5 et 9 ans, idem pour le nombre de personnes dans ces catégories. Des intervalles moins grands ont aussi été considérés. Enfin, une dernière catégorie d'intervalles a été testée. Celle-ci a été construite grâce à un arbre de décision utilisé pour déterminer les âges seuils pour lesquels les comportements de déplacement changent. Toutes les possibilités citées ci-dessus ne se sont pas révélées très pertinentes, soit parce que les seuils d'âges varient trop d'une année d'enquête à une autre, soit parce que le nombre d'intervalles est beaucoup trop grand. Il est donc intéressant de se pencher sur d'autres variables pour représenter la distribution des âges dans un ménage. Celles-ci refléteraient idéalement les différentes générations, l'écart d'âge entre les générations, etc. Des variables descriptives comme l'âge moyen, l'écart type des âges, l'amplitude maximale des âges, le minimum et le maximum ont été considérées. L'âge moyen du ménage et l'amplitude maximale des âges apportent tous deux des informations complémentaires. Ainsi, la combinaison des deux est retenue.

L'amplitude maximale des âges permet un groupement des ménages suivant les générations, tandis que l'âge moyen des ménages apporte des informations sur l'âge de ceux-ci. Sur la Figure 4.1, il est possible de distinguer différents types de ménages. Premièrement les ménages dont l'amplitude maximale des âges est faible correspondent à des ménages à une seule génération. Au sein de ce groupe, les « aînés » (en haut à gauche) se distinguent des plus jeunes (en bas à gauche). Les « aînés » sont souvent des personnes retraitées ou des couples dont les enfants ont quitté le foyer. Les plus jeunes, eux, représentent souvent des colocations de vingtenaires, des couples ou des célibataires. La plupart des ménages composés de deux générations sont repérables au milieu du graphique; ces ménages sont souvent des adultes avec enfants. Dans cette même catégorie, un autre type de ménage est également observé, même s'ils sont moins nombreux : ce sont les adultes avec personnes âgées, dont l'âge moyen est plus élevé. Enfin,

quelques observations ont une amplitude maximale des âges supérieure à 50 ans : celles-ci correspondent aux ménages à trois générations (un enfant, un adulte et une personne âgée au moins) ou aux ménages avec un saut de génération (un enfant et une personne âgées) ou bien à des ménages dont l'écart entre les générations est très grand (couple ayant eu des enfants tard...).

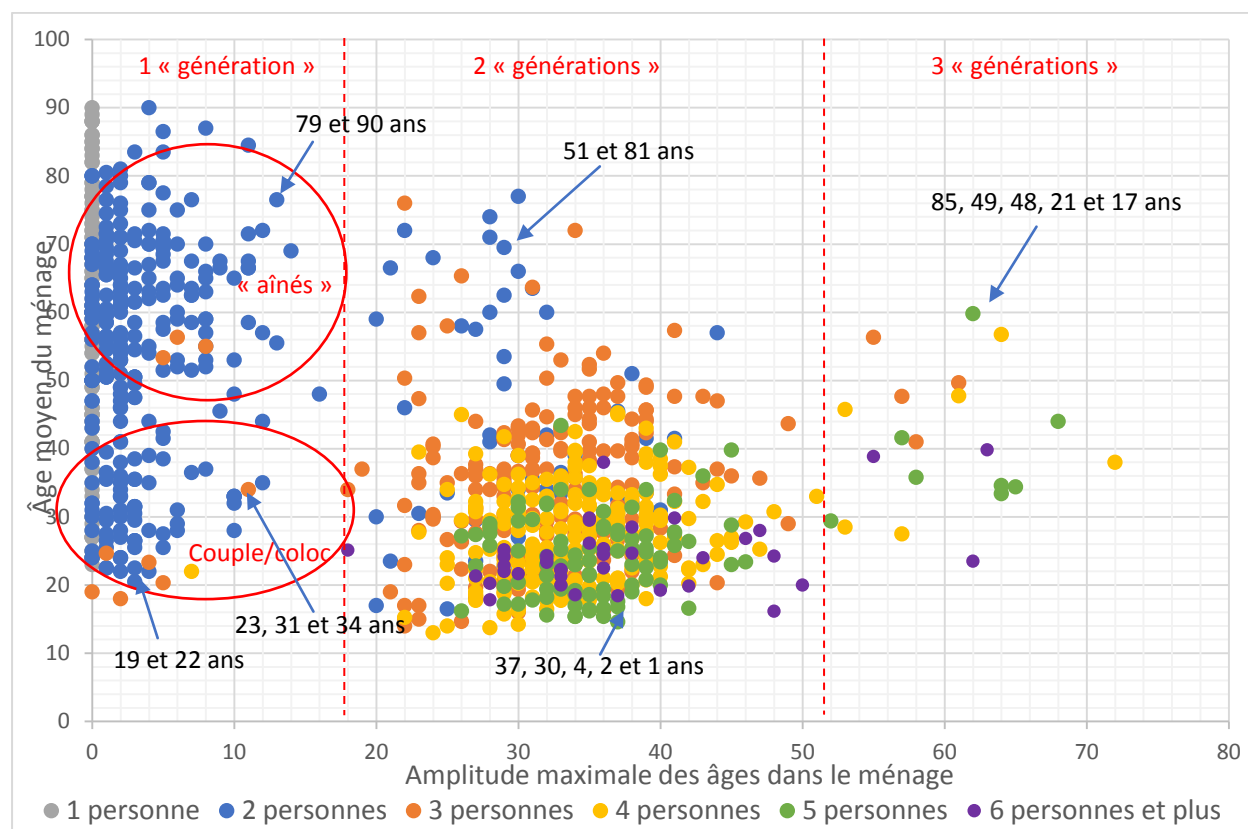


Figure 4.1 : Répartition des ménages suivant leur taille, leur âge moyen et l'amplitude maximale des âges dans le ménage

La Figure 4.2 montre, quant à elle, le nombre de ménages suivant l'amplitude maximale des âges et la taille du ménage. Ainsi, peu de ménages ont une amplitude maximale des âges située entre 12 et 20 ans ou supérieure à 45 ans. Il est également observé que la quasi-totalité des ménages dont l'amplitude des âges est inférieure à 20 ans sont des ménages de deux ou trois personnes. Il est à noter que les ménages individuels ne sont pas représentés sur ce graphique, puisque l'intégralité des ménages d'une personne ont une amplitude maximale des âges de 0.

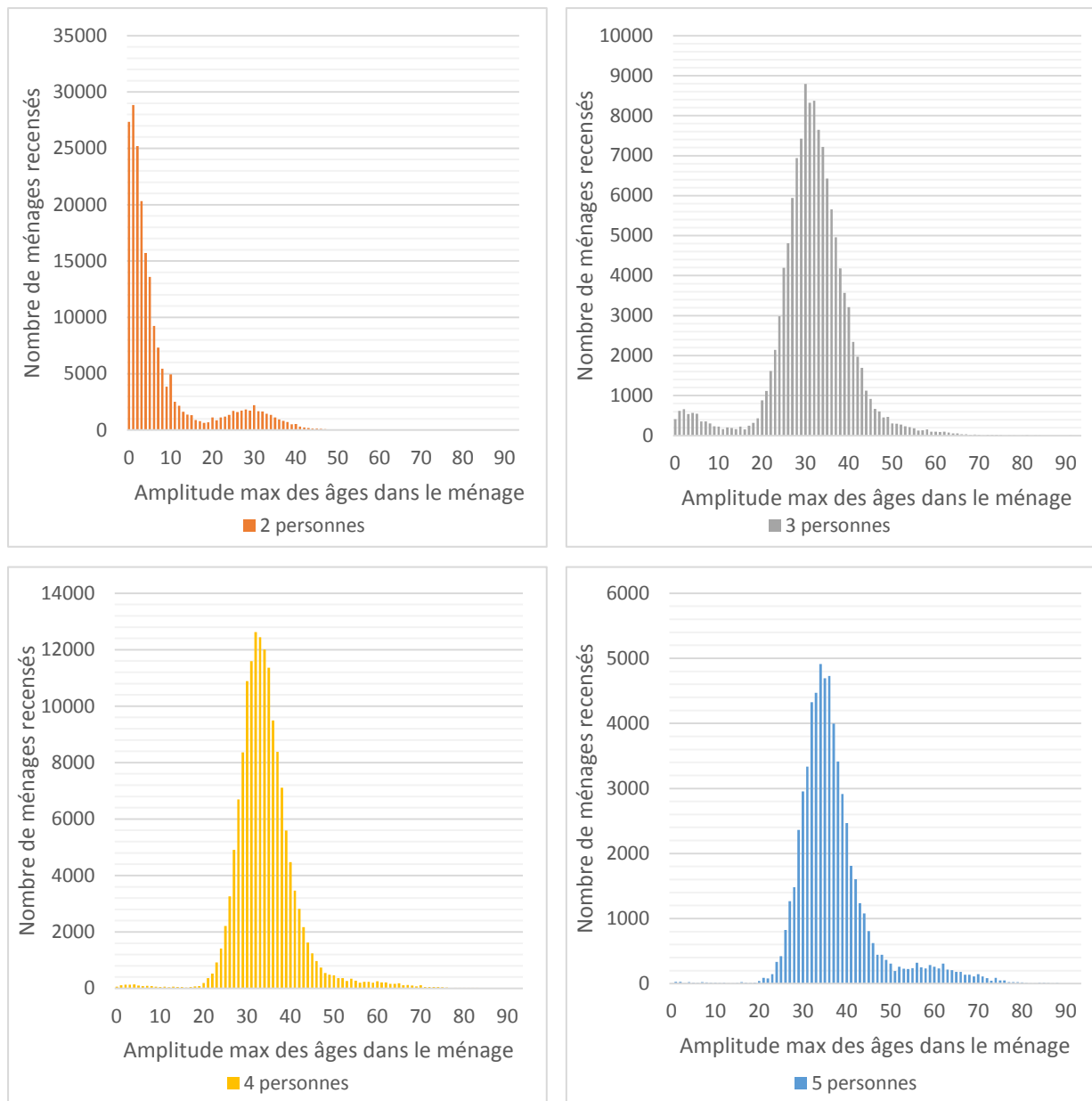


Figure 4.2 : Nombre de ménages en fonction de l'amplitude maximale des âges dans le ménage et de la taille de celui-ci

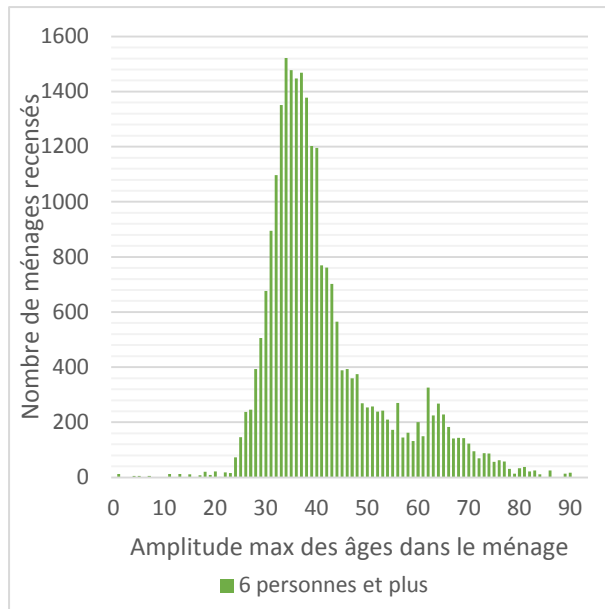


Figure 4.3 : Nombre de ménages en fonction de l’amplitude maximum des âges dans le ménage et de la taille de celui-ci (suite)

Enfin la Figure 4.4 représente le nombre de ménages suivant l’âge moyen du ménage et sa taille. Ici aussi, plusieurs observations peuvent être faites. Premièrement, l’âge moyen des ménages se situe globalement entre 10 et 85 ans avec deux « pics » : autour de 25 ans et autour de 60 ans (ce dernier étant bien sûr beaucoup moins important). La quasi-totalité du pic autour de 25 ans est due aux ménages de quatre personnes et plus. Deuxièmement, il faut remarquer que plus la taille du ménage est grande plus l’âge moyen est faible. Enfin, la somme des ménages de deux et trois personnes est à peu près constante sur des âges moyens allant de 18 à 65 ans, mais les ménages de trois personnes sont plus nombreux lorsque l’âge moyen est plus faible, tandis que le contraire est observé pour les ménages de deux personnes. Ceci représente probablement en majorité les couples avec un enfant ou sans enfant et la transition au départ de l’enfant du ménage.

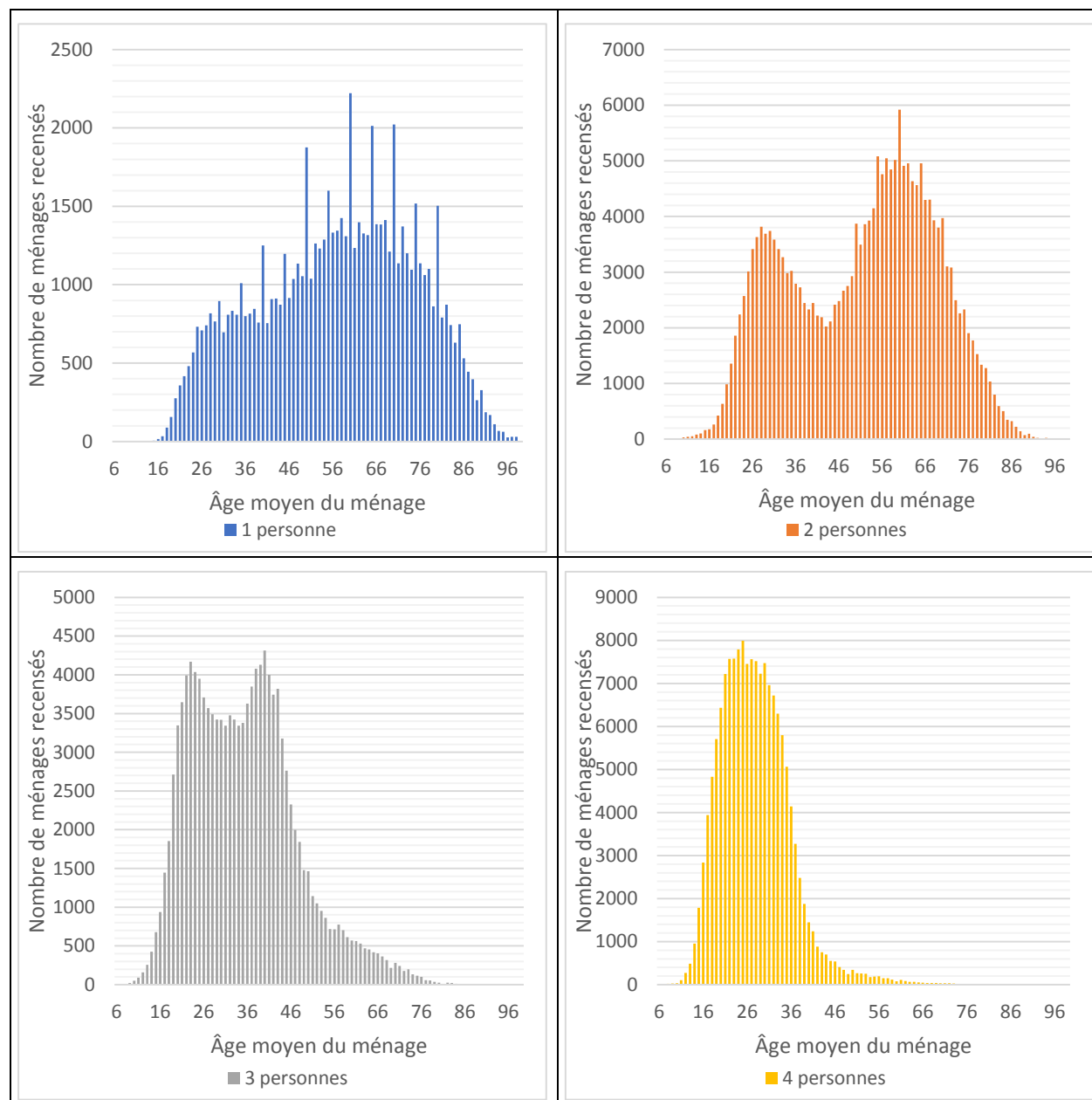


Figure 4.4 : Nombre de ménages en fonction de l'âge moyen du ménage et de la taille de celui-ci

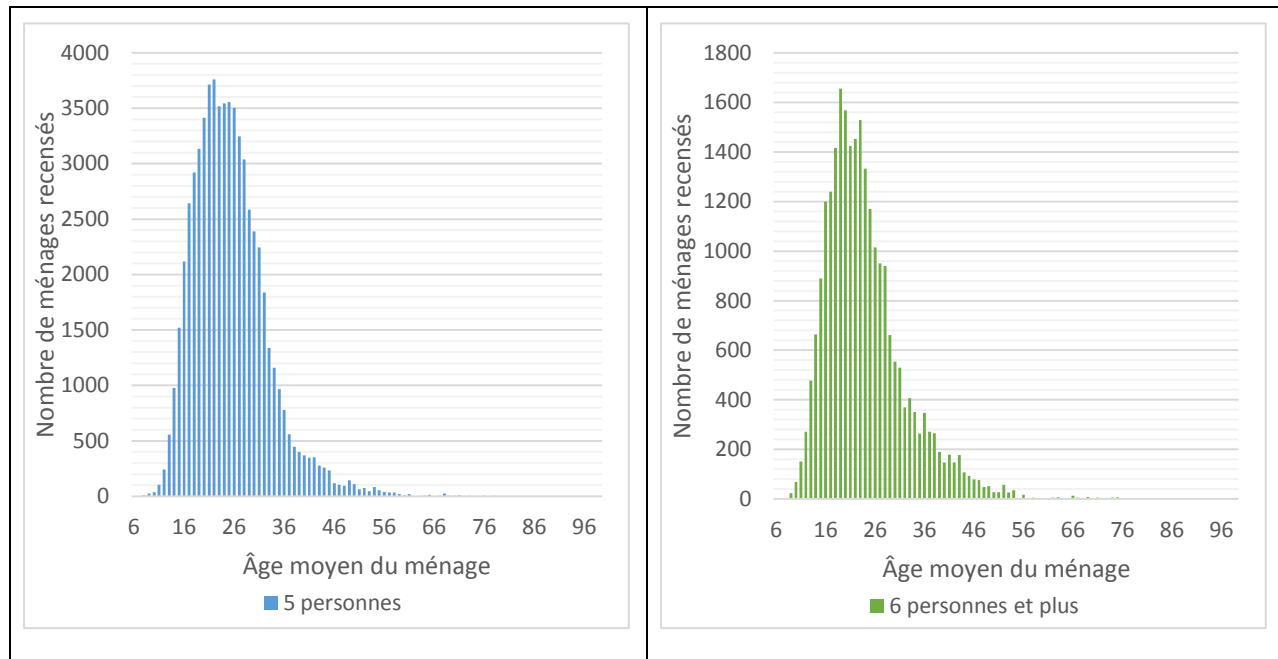


Figure 4.5 : Nombre de ménage en fonction de l'âge moyen du ménage et de la taille de celui-ci (suite)

4.2 Sélection des variables pertinentes pour expliquer les comportements de déplacement

La section de résultats suivante se décompose en deux parties. Premièrement, les résultats des arbres de décision utilisés pour déterminer quelles variables du ménage ont une influence sur les déplacements de la personne sont présentés. Ensuite, l'étude des différences de déplacements qu'entraîne la variabilité des caractéristiques du ménage révélées pertinentes sera faite suivant deux processus différents.

4.2.1 Arbres de décision

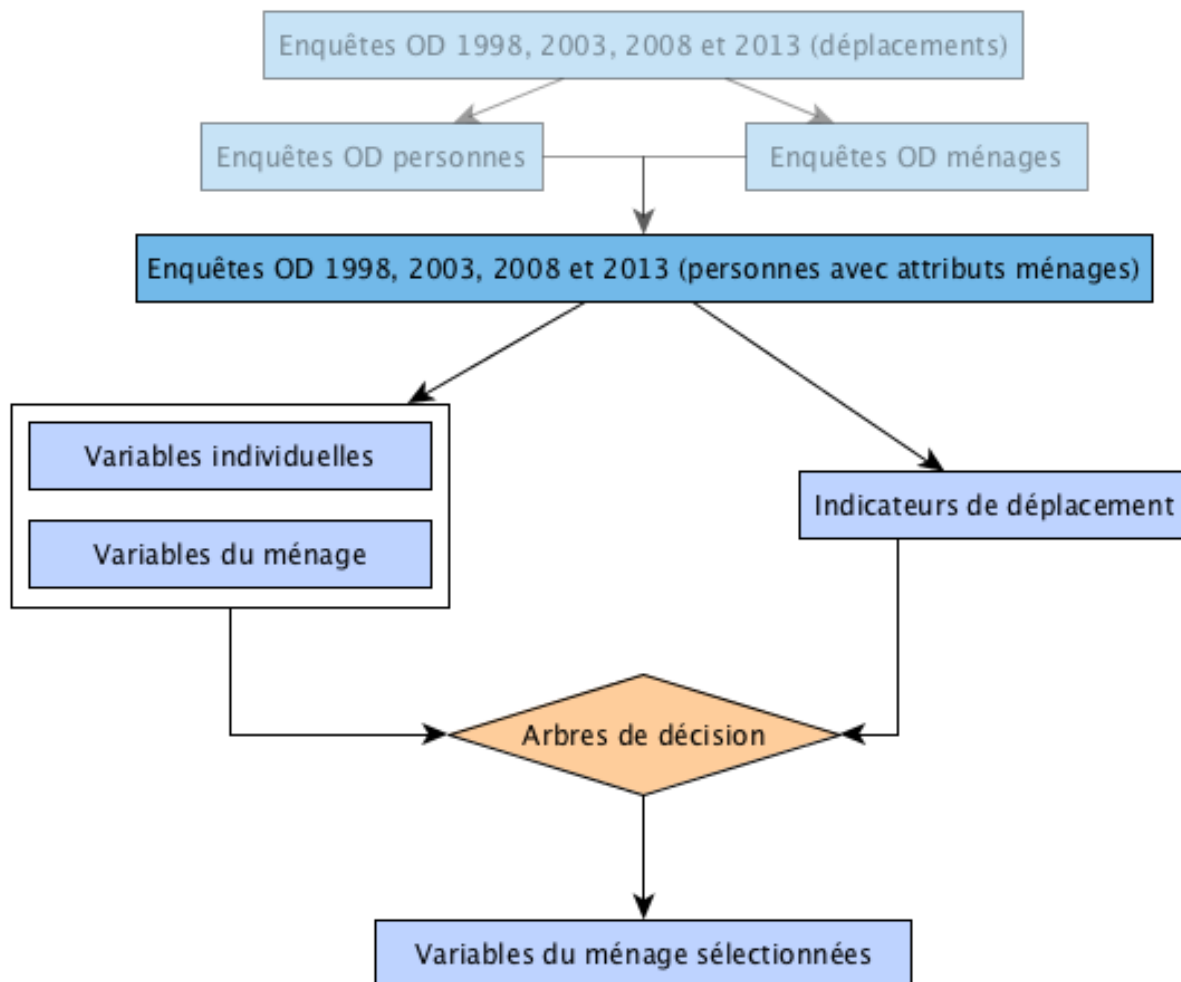


Figure 4.6 : Position de la section suivante dans le schéma méthodologique

Pour sélectionner les variables de la personne et du ménage ayant une influence sur les comportements de déplacement individuels, des arbres de décision sont utilisés. Un arbre est créé pour chacun des cinq indicateurs de déplacement sélectionnés, en prenant en compte les 13 variables mentionnées dans la section 3.3.2 de la méthodologie, ceci pour chacune des quatre années d'enquête. Vingt arbres de décision sont donc obtenus. Les résultats sont détaillés pour les arbres de l'enquête 2013, la plus récente.

Pour l'enquête 2013, le nombre d'observations dans la base de données personnes (une fois les 0-5 ans enlevés) est de 178 908. La taille minimum des feuilles utilisées pour la création des cinq arbres est 1500 (soit environ 1% des observations pour chacune des feuilles). Ce nombre a pu

être modifié suivant l'indicateur en jeu afin de ne pas obtenir d'arbres avec trop peu de nœuds ou au contraire, beaucoup trop de nœuds qui rendraient l'arbre illisible. Le paramètre de complexité lui, varie suivant l'indicateur représenté, il est plus petit pour des indicateurs pour lesquels une des éventualités est sous représentée (le fait de faire une chaîne complexe ou pas par exemple). Les résultats pour les cinq indicateurs suivant l'enquête OD 2013 sont donnés ci-dessous. Pour les années 1998, 2003 et 2008, la taille minimum des feuilles et le paramètre de complexité varient. Bien que les arbres obtenus présentent un taux d'erreur plus ou moins bon suivant l'enquête et l'indicateur représenté, les résultats sont similaires à ceux de 2013. Les arbres de décision obtenus pour ces années ne seront donc pas détaillés dans ce mémoire. Cependant, ils ont bien sûr servi au choix des variables (avec le même poids que ceux de 2013) et seront présents dans l'étude de l'importance des variables en section 4.2.2.

4.2.1.1 Être mobile ou pas

Le premier indicateur : « Être mobile ou pas » témoigne de la mobilité ou de l'immobilité des personnes sur la journée d'enquête. Une personne ayant effectué un déplacement ou plus sera considérée comme mobile dans l'enquête tandis qu'une personne qui ne s'est pas déplacée sera considérée comme non mobile. Dans l'enquête OD 2013, environ 18% de la population est recensée comme non mobile. L'arbre de décision permet de séparer la population suivant les caractéristiques personnelles ou des ménages et de prédire pour chaque groupe ainsi créé si les personnes sont mobiles ou non mobiles. Peu de groupes sont prédits comme non mobiles. L'entière des moins de 60 ans sont prédits comme mobiles. Parmi les 60 ans et plus, seuls quelques groupes sont prédits non mobiles. Chez les personnes âgées de 77 ans ou plus, celles ayant un accès à l'automobile inférieur à 0,47, celles ayant un accès à l'auto supérieur ou égal à 0,47 et une amplitude maximale des âges dans le ménage supérieure à 18 ans, ou bien inférieure à 18 ans, mais avec un âge moyen supérieur ou égal à 85 ans, sont prédites non mobiles. Chez les moins de 77 ans, seuls les individus ayant un accès à l'auto inférieur à 0,46, étant âgés de 65 ans ou plus et vivant dans les régions de l'Est, de Laval, de Longueuil et des couronnes Nord et Sud sont prédits comme non mobiles.

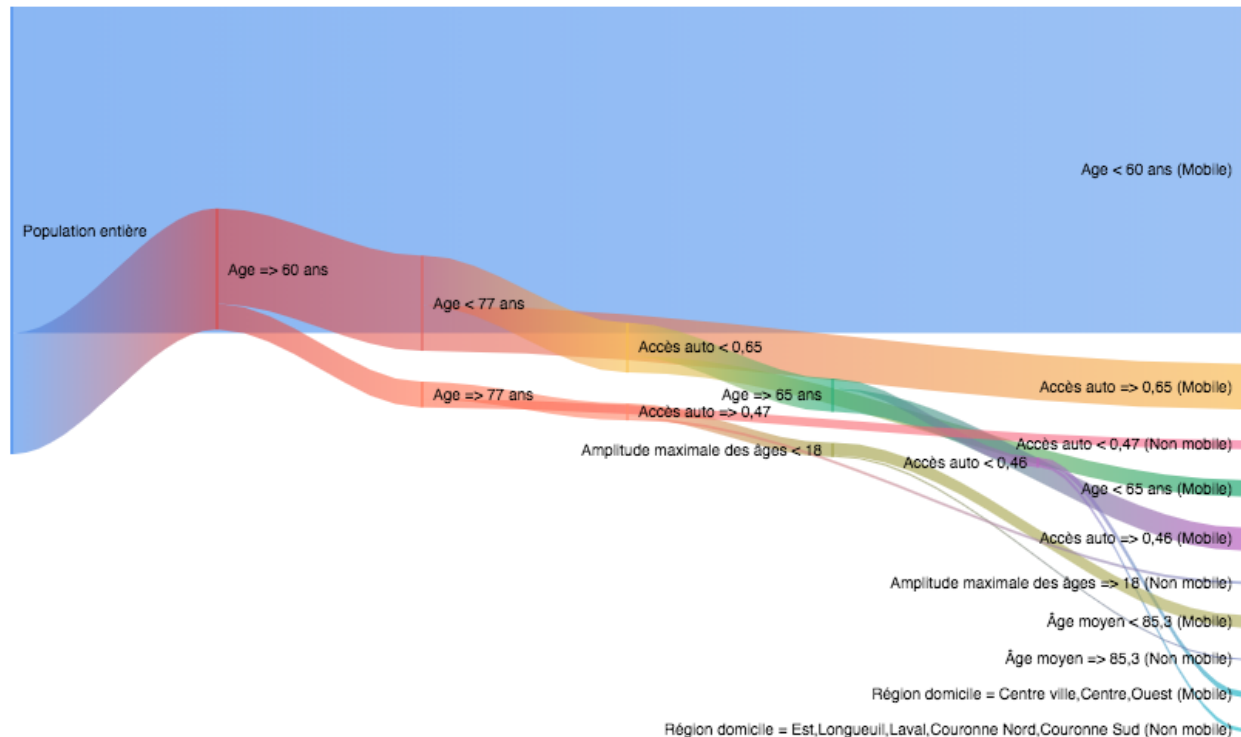


Figure 4.7 : Arbres de décision "Être mobile ou pas", enquête OD 2013

Le taux d'erreur de l'algorithme est de 0,1715 c'est-à-dire qu'environ 17% des personnes observées ont été prédites comme mobiles alors qu'elles ne le sont pas dans l'enquête OD 2013 (ou inversement). La Figure 4.8 et le Tableau 4.1 représentent la matrice de confusion de l'arbre de décision représenté en Figure 4.7. Ainsi, il est possible de voir que la majorité du taux d'erreur est dû à des personnes observées non mobiles dans l'enquête OD et prédites mobiles par l'algorithme. Il y a donc une sous-représentation des non mobiles par l'algorithme. Ceci est dû au fait que le nombre d'observations est bien plus faible pour les non mobiles, l'algorithme a donc plus de mal à les prédire. Il est aussi possible de supposer qu'aucune variable incluse dans le processus n'est suffisamment discriminante pour cette propriété.

Tableau 4.1 : Matrice de confusion de l'arbre "Être mobile ou pas", enquête OD 2013

	Mobile (« OUI »)	Non mobile (« NON »)	Total prédits
Mobile (« OUI »)	143 933	28 103	172 036
Non mobile (« NON »)	2 575	4 297	6 872
Total observés	146 508	32 400	178 908

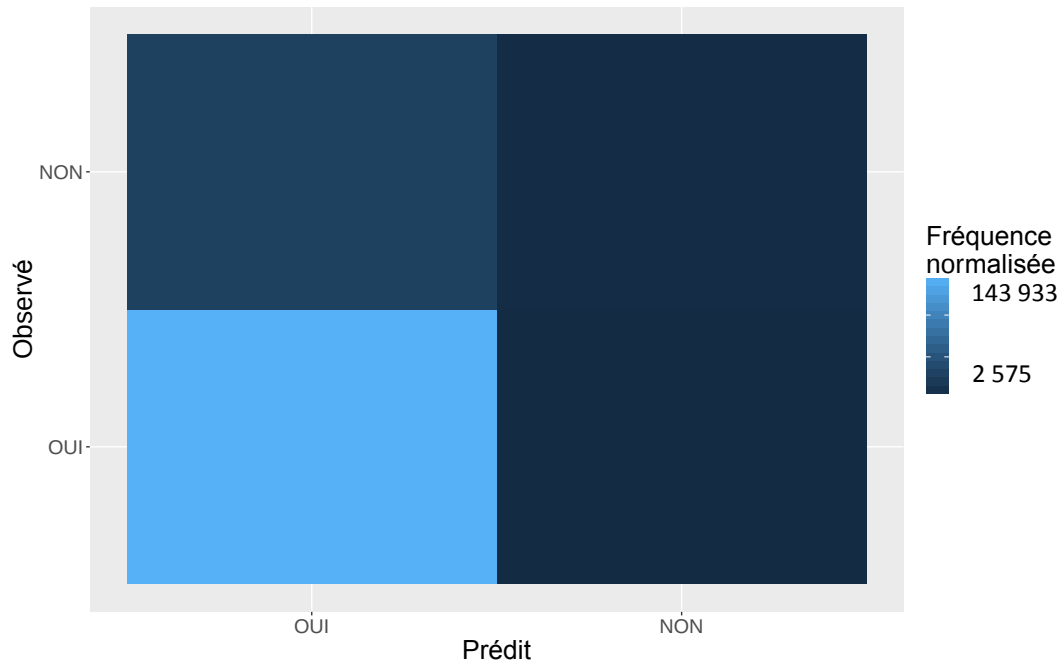


Figure 4.8 : Matrice de confusion de l'arbre "Être mobile ou pas", enquête OD 2013

4.2.1.2 Nombre de déplacements

En ce qui concerne le nombre de déplacements effectués par personne sur une journée d'enquête (déplacements « retour » non comptés), l'arbre ne permet pas de prédire un nombre de déplacements de deux ou plus (Figure 4.9). Plusieurs arbres de décision (un pour chaque nombre de déplacements de 1 à 6 et plus) ont donc été exécutés. Pour ces derniers, l'arbre permet de dire pour chaque groupe de personnes séparées selon leurs caractéristiques personnelles et du ménage, si elles font oui ou non x déplacements. Pour ces arbres, la taille minimum des feuilles utilisée est plus petite, car le nombre d'observations (surtout pour les grands nombres de déplacements) est plus faible. Un exemple pour trois déplacements est donné en Figure 4.10. La Figure 4.11 présente la matrice de confusion de ce dernier. Ici aussi, le fait de faire un grand nombre de déplacements est dur à représenter avec les arbres de décision du fait du petit nombre d'observations. Ainsi, quasiment la totalité des personnes observées dans l'enquête OD comme ne faisant pas trois déplacements sont effectivement prédites dans la même catégorie par l'arbre de décision, mais au contraire, les personnes faisant réellement trois déplacements sont très peu prédites comme telles par l'algorithme. Cependant, le nombre de personnes effectuant trois déplacements est très faible, le taux d'erreur de cet arbre reste donc faible (7%).

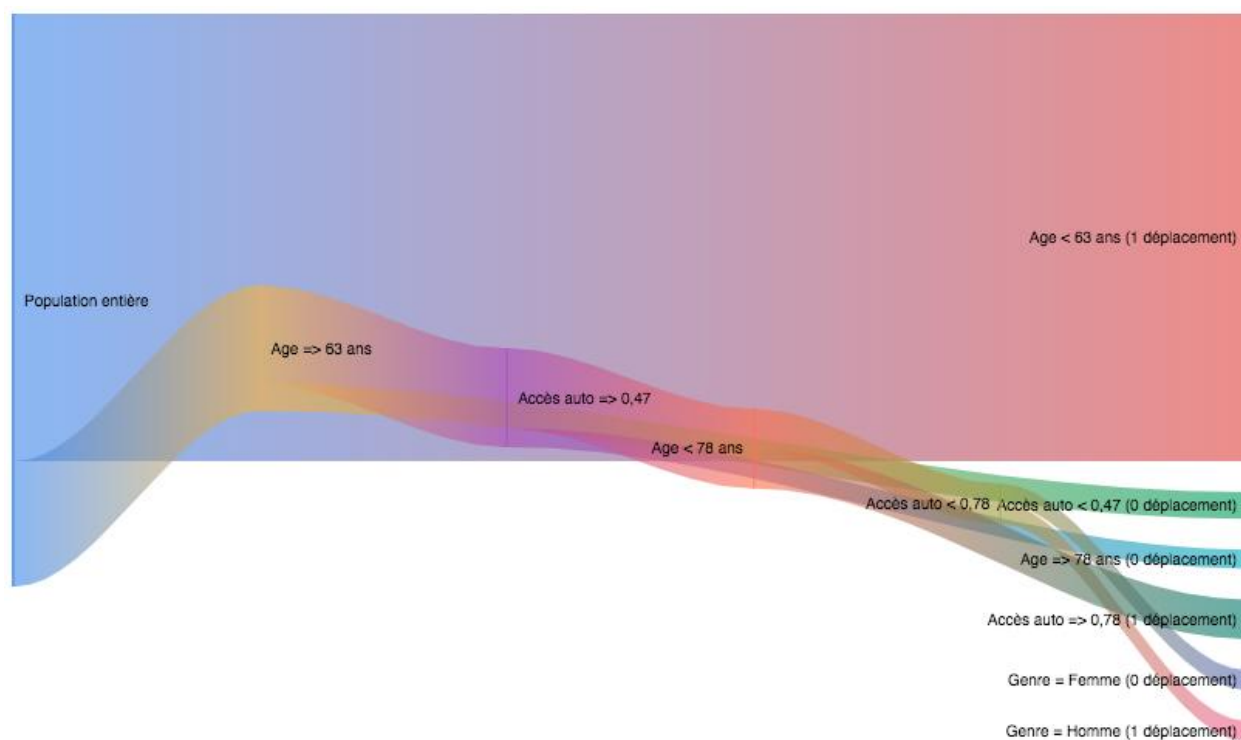


Figure 4.9 : Arbre de décision "Nombre de déplacements", enquête OD 2013

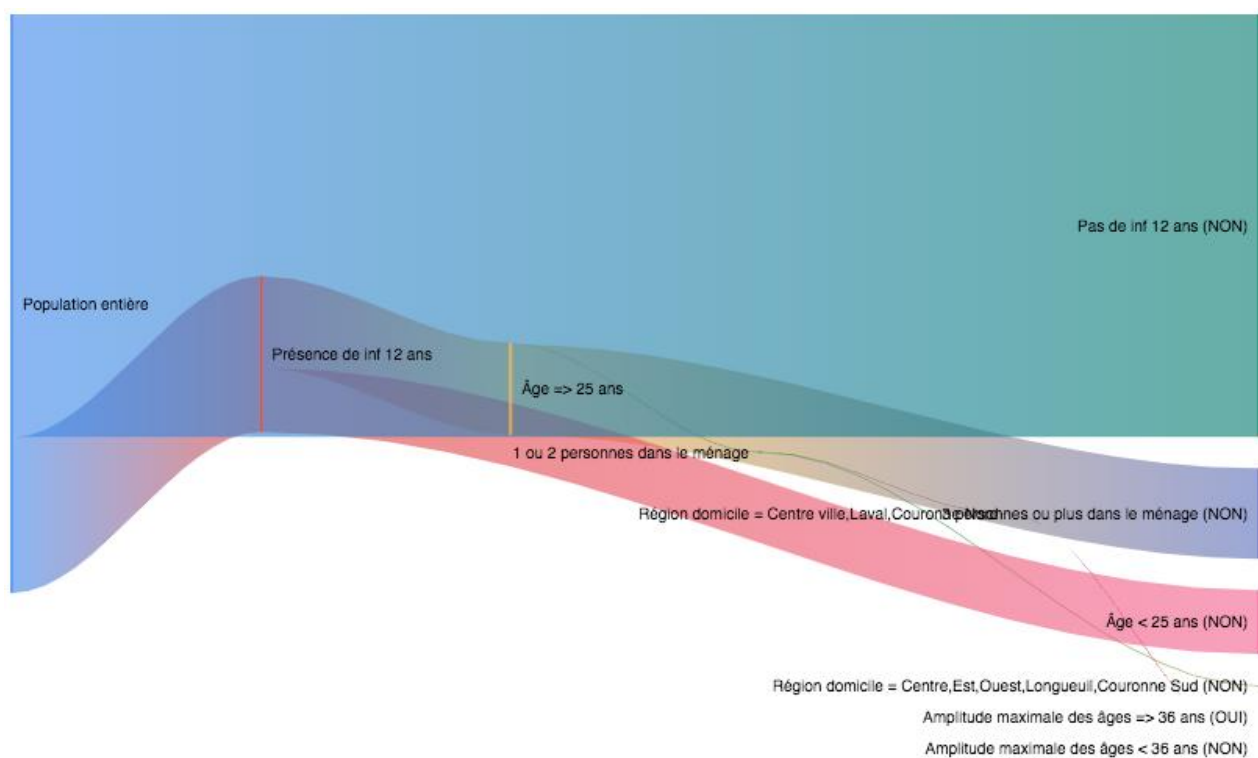


Figure 4.10 : Arbre de décision "Faire 3 déplacements", enquête OD 2013

Tableau 4.2 : Matrice de confusion de l'arbre "Faire 3 déplacements", enquête OD 2013

	3 déplacements (« OUI »)	Autre (« NON »)	Total prédits
3 déplacements (« OUI »)	13	12	25
Autre (« NON »)	12 579	166 304	178 883
Total observés	12 592	166 316	178 908

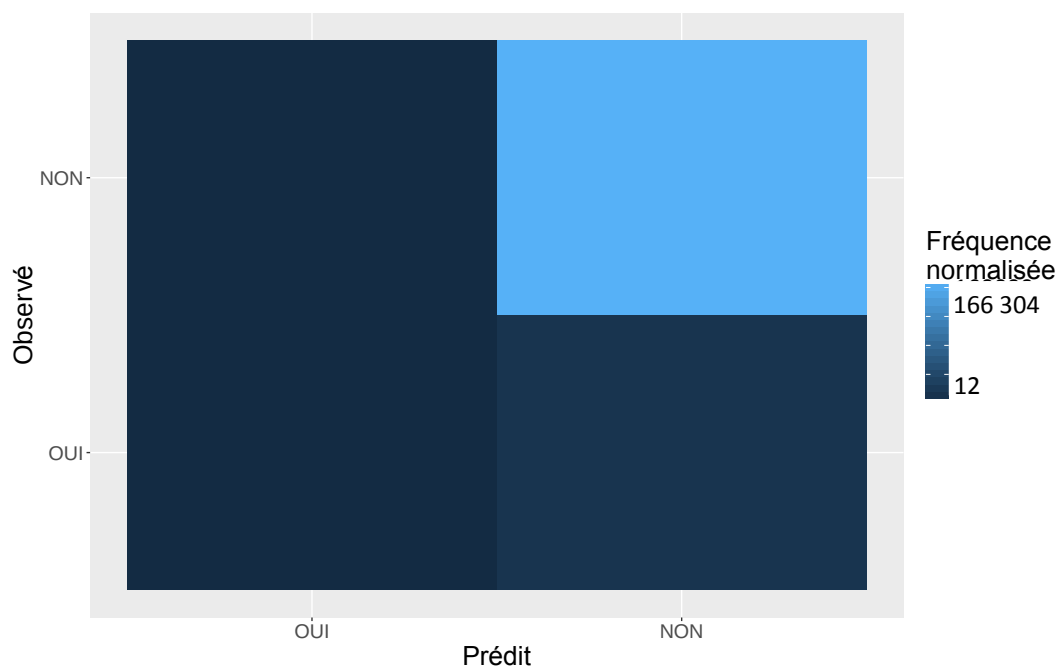


Figure 4.11 : Matrice de confusion de l'arbre "Faire 3 déplacements", enquête OD 2013

4.2.1.3 Déplacements en pointe AM

Pour la part de déplacements en pointe matinale, l'arbre de décision prédit une valeur de la part des déplacements effectués entre 6h et 9h le matin. Pour cette étude, seules les personnes mobiles sont considérées. Sur la population entière de l'enquête OD 2013, les moins de 17 ans sont ceux qui se voient attribuer la part de déplacements en pointe matinale la plus élevée (0,44 pour les moins de 12 ans et 0,40 pour les 12-17 ans), suivis des 34-59 ans qui vivent avec des enfants de moins de 12 ans (0,32). Au contraire, la part de déplacements en pointe matinale des plus de 65 ans dont l'âge moyen du ménage est supérieur à 70 ans est de 0,06.

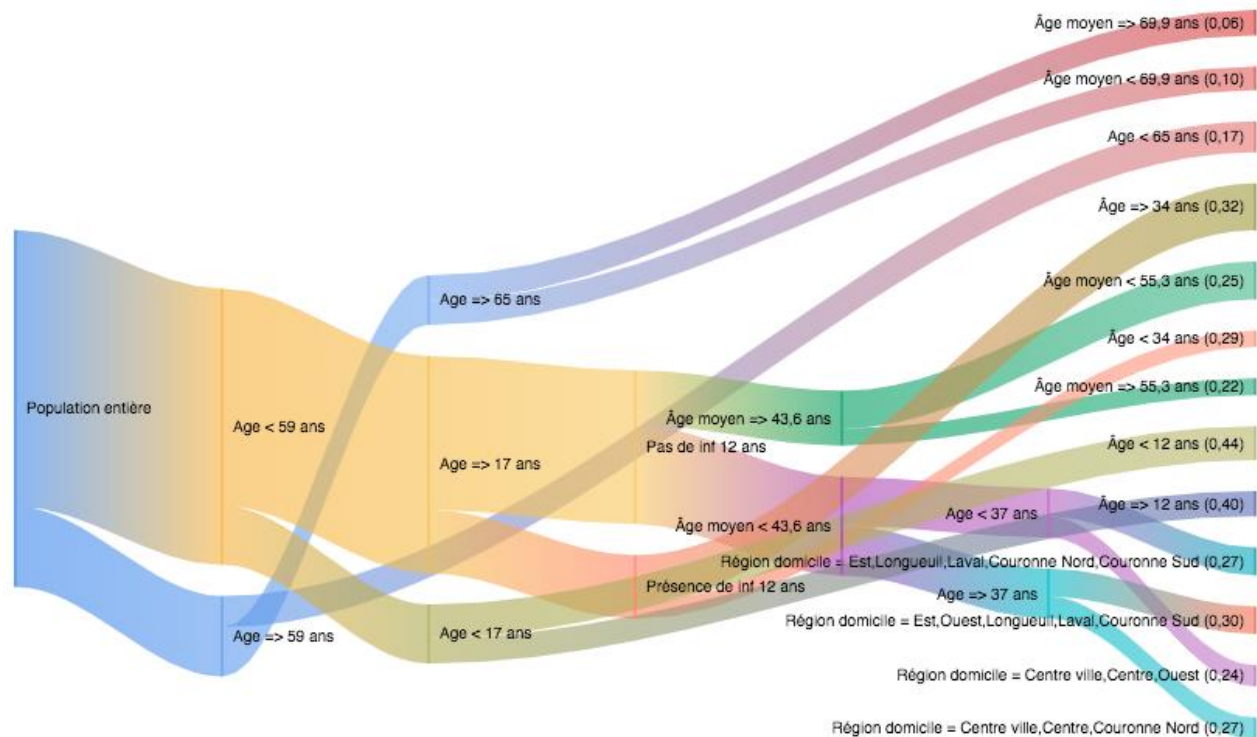


Figure 4.12 : Arbre de décision "Part de déplacements en pointe AM", enquête OD 2013

Pour un arbre de régression (qui prédit une valeur continue), il est difficile de représenter une matrice de confusion. Ainsi des catégories ont été créées pour la valeur observée et la valeur prédite de la part de déplacement en pointe matinale et la matrice de confusion a été élaborée avec ces valeurs. Elle est représentée en Figure 4.13 et dans le Tableau 4.3. Très peu de parts de déplacement en pointe matinale de 0,6 et plus sont observées, l'arbre, lui, n'en prédit aucune. Pour les parts plus faibles, l'arbre prédit trop de personnes dans l'intervalle $[0,2; 0,4)$. À titre comparatif, des distributions aléatoires ont été testées, les taux d'erreur pour ces dernières restent supérieurs à celui de la distribution prédite par l'arbre, qui est de 49% ce qui est tout de même très important.

Tableau 4.3 : Matrice de confusion de l'arbre "Part de déplacements en pointe AM", enquête OD 2013

	[0; 0,2)	[0,2; 0,4)	[0,4; 0,6)	[0,6; 0,8)	[0,8; 1]	Total prédits
[0; 0,2)	24 812	2 611	4 917	202	499	33 041
[0,2; 0,4)	35 844	15 347	33 836	2 493	1 689	89 209
[0,4; 0,6)	1 995	4 267	17 770	138	88	24 258
[0,6; 0,8)	0	0	0	0	0	0
[0,8; 1]	0	0	0	0	0	0
Total observés	62 651	22 225	56 523	2 833	2 276	146 508

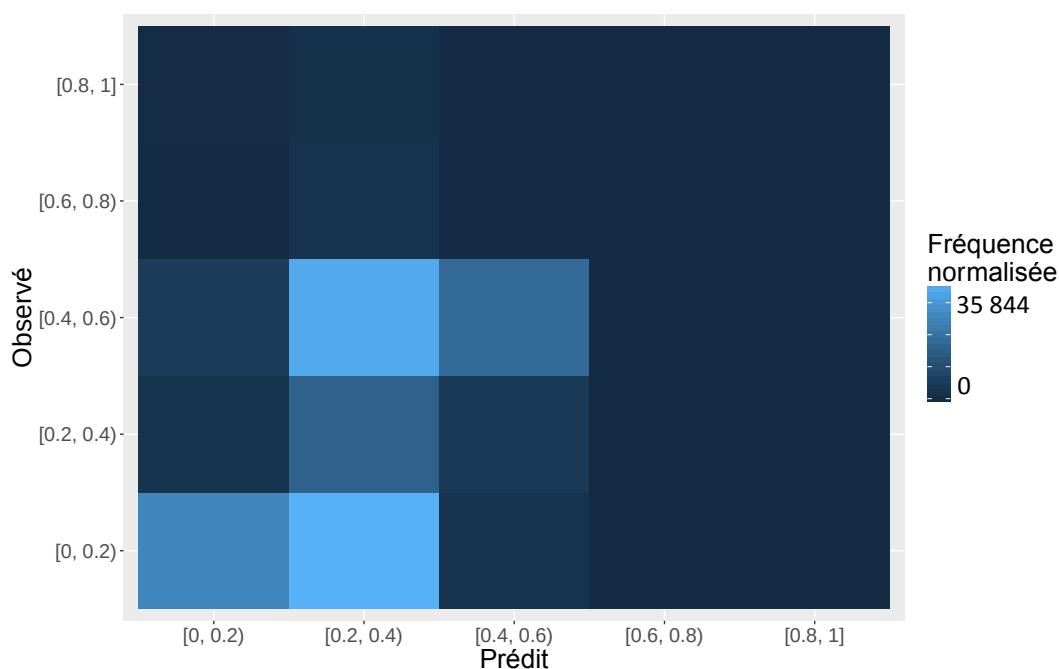


Figure 4.13 : Matrice de confusion de l'arbre "Part de déplacements en pointe AM", enquête OD 2013

4.2.1.4 Déplacement en automobile

L'arbre de décision pour la part de déplacements en automobile (conducteur ou passager) prédit également un nombre entre 0 et 1. Ici aussi, seules les personnes mobiles sont considérées. Les plus de 21 ans habitant à Laval ou dans les couronnes Nord et Sud et dont l'accès à l'auto est supérieur à 0,38 ont une part de déplacements en automobile record (0,94). Hormis les personnes

dont l'accès à l'auto est très faible (inférieur à 0,1), les jeunes de moins de 27 ans sont ceux qui utilisent le moins leur automobile.

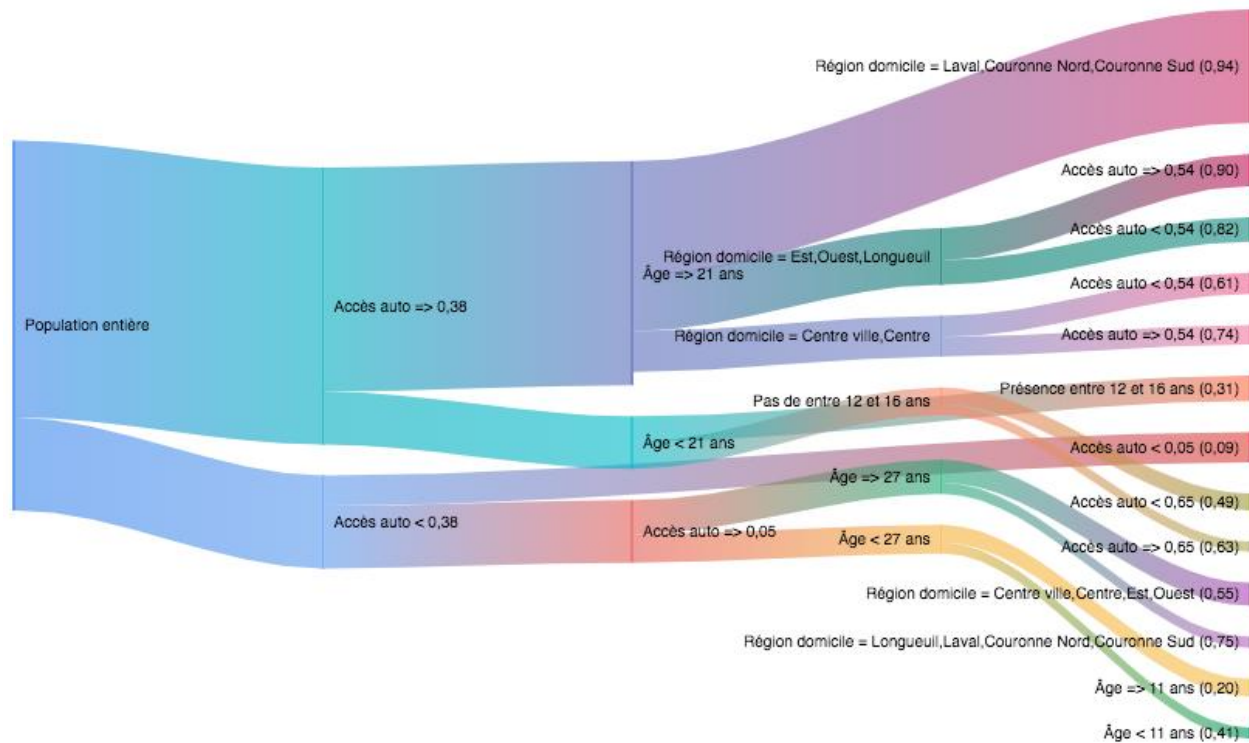


Figure 4.14 : Arbre de décision "Part de déplacements en auto", enquête OD 2013

De même que pour la part de déplacement en pointe matinale, la matrice de confusion a ici aussi été effectuée avec des catégories de part. Elle est représentée en Figure 4.15. L'arbre effectue un très bon travail pour les personnes ayant une part de déplacement en auto observée dans l'intervalle $[0,8; 1]$, mais il sous prédit les personnes ayant une part de déplacement en auto faible. Le taux d'erreur est de 60% environ. Ici aussi, des distributions aléatoires ont été appliquées, un taux d'erreur supérieur à 90% est donné pour les 10 itérations effectuées. Ainsi, même si le taux d'erreur de la distribution prédite par l'arbre de régression est important, il reste plus faible que celui d'une distribution aléatoire.

Tableau 4.4 : Matrice de confusion de l'arbre "Part de déplacements en automobile", enquête OD 2013

	[0; 0,2)	[0,2; 0,4)	[0,4; 0,6)	[0,6; 0,8)	[0,8; 1]	Total prédictions
[0; 0,2)	10 565	218	413	102	729	12 027
[0,2; 0,4)	10 802	435	2 412	549	2 858	17 056
[0,4; 0,6)	8 293	428	2 358	730	8 251	20 060
[0,6; 0,8)	7 262	487	2 434	999	18 770	29 952
[0,8; 1]	4 008	268	2 447	1 298	59 392	67 413
Total observés	40 930	1 836	10 064	3 678	90 000	146 508

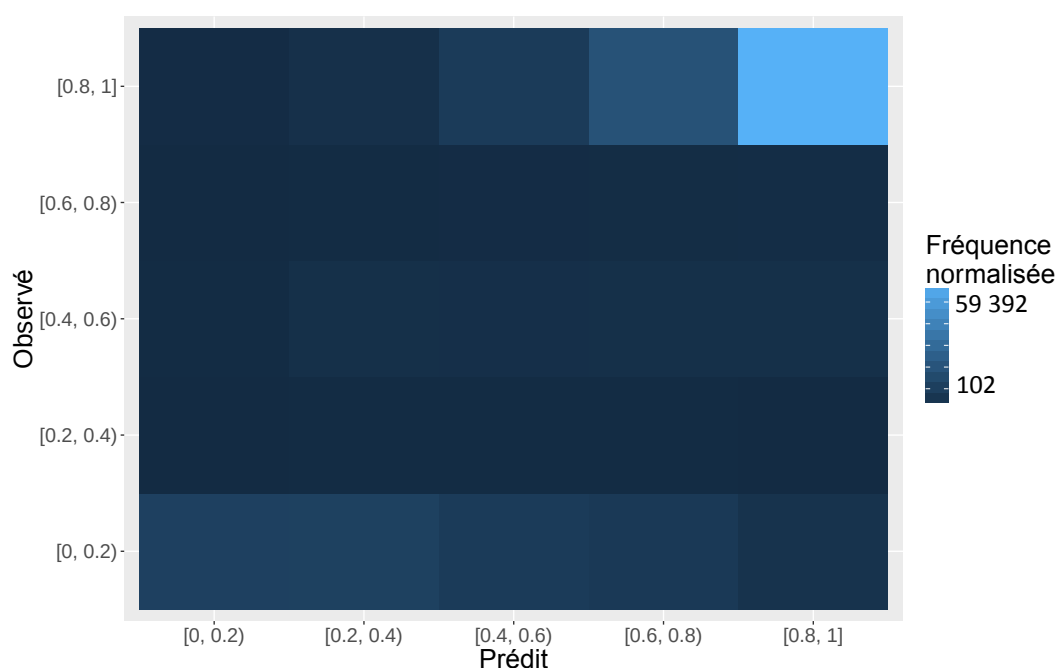


Figure 4.15 : Matrice de confusion de l'arbre "Part de déplacements en automobile", enquête OD 2013

4.2.1.5 Faire une chaîne complexe ou non

Pour le fait de faire une chaîne complexe ou non, l'arbre (Figure 4.16) est difficilement interprétable, car le nombre de personnes pour lesquelles ce dernier prédit une chaîne complexe ou plus dans la journée d'enquête est très faible (moins de 1% de la population).

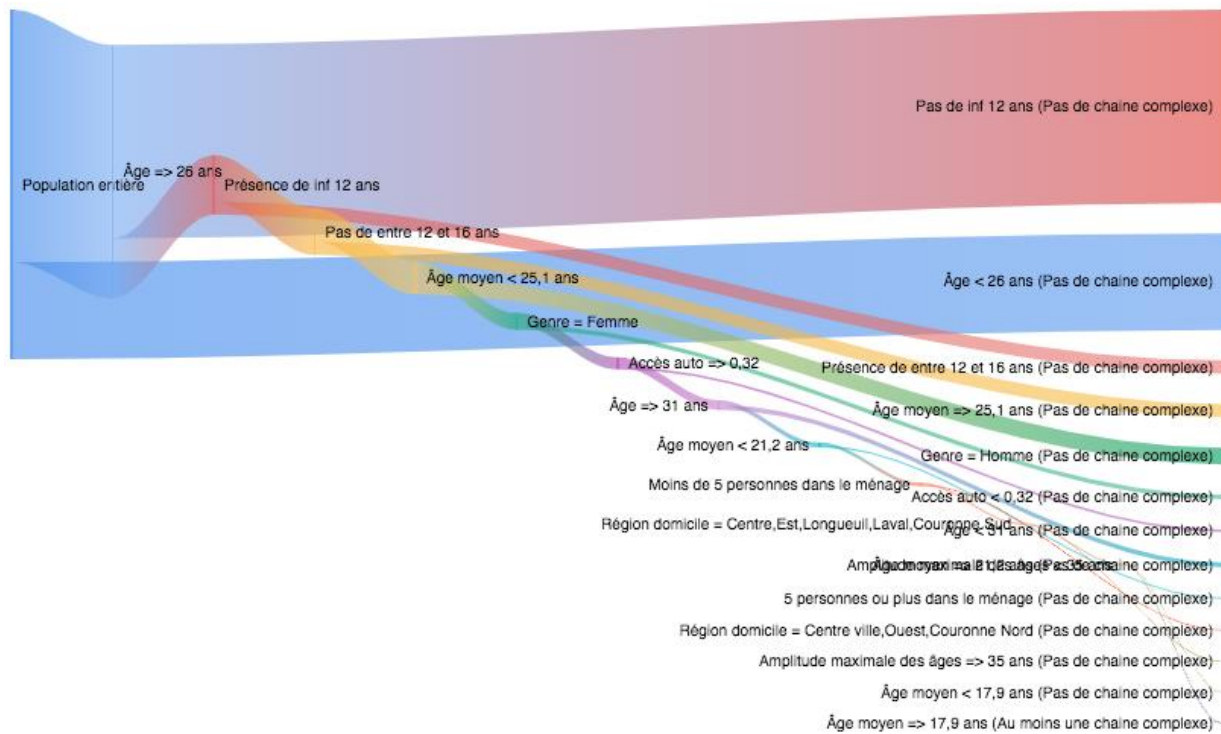


Figure 4.16 : Arbre de décision "Faire une chaîne complexe ou pas", enquête OD 2013

Le taux d'erreur de la plupart des arbres de décision ou de régression effectués est élevé, ce qui permet de conclure que les arbres ne sont pas très performants pour prédire la valeur ou l'intervalle des indicateurs de déplacement. Ceci pourrait être un frein à leur utilisation. Cependant, ils ne sont pas utilisés à ces fins dans cette étude.

4.2.2 Importance des variables et processus de sélection

Le but n'est pas de regarder quelles valeurs sont prédites pour chaque indicateur, mais plutôt de se pencher sur les variables qui ont une influence sur ces indicateurs. Il faut donc regarder les variables qui influent sur le processus de création des branches des arbres. Comme cela a été précisé dans la méthodologie (section 3.4), il ne suffit pas de regarder les variables aux nœuds pour dire que ce sont les plus importantes dans l'arbre de décision. Il faut également regarder l'importance des variables donnée par l'algorithme. Les Tableau 4.5, Tableau 4.6, Tableau 4.7 et Tableau 4.8 présentent l'importance des différentes variables pour chaque indicateur pour les années 1998 à 2013. Pour chaque indicateur (en colonne), l'importance relative de chaque variable est donnée (plus le ton est foncé, plus la variable est importante pour l'indicateur

concerné). La dernière colonne représente la somme des importances pour tous les indicateurs de chaque variable. Ici aussi, plus le ton est foncé, plus la variable est importante.

Tableau 4.5 : Importance des variables pour l'année 1998

VARIABLE/INDICATEUR	Part auto	Part pointe am	Chaîne complexe	Mobile	1 dep	2 dep	3 dep	4 dep	5 dep	6 et plus dep	Somme
ACCES_AUTO	5927,4	34,9	339,3	579,5	53,1	26,8	16,7	6,8	5,2	4,9	6994,7
AGE	5513,3	890,5	1355,1	2811,2	1816,6	227,7	67,2	38,5	16,2	11,6	12748,0
AGE_AMP_MAX	99,2	31,7	906,3	61,3	108,9	29,0	14,7	11,5	12,0	8,1	1282,7
AGE_MEAN	211,6	426,2	1332,2	2033,9	502,7	170,3	58,8	24,5	15,8	19,2	4795,3
NB_FEMMES_MENAGE	80,0	23,1	293,2	27,2	64,7	7,9	8,5	5,5	4,3	4,9	519,3
NB_HOMMES_MENAGE	74,6	19,3	314,2	121,7	53,2	6,6	6,8	4,9	5,2	5,3	611,9
NB_PERSONNES_MENAGE	279,9	30,5	360,2	116,3	185,6	45,1	31,5	17,1	6,1	4,7	1077,0
pres_ENTRE12_16ANS	329,4	1,6	133,9	0,0	49,2	NA	2,1	2,7	0,8	0,3	520,2
pres_ENTRE76_84ANS	0,2	31,3	33,9	375,9	77,0	20,8	0,2	0,9	0,0	0,1	540,5
pres_INF_12ANS	122,8	17,0	103,8	NA	142,9	NA	74,2	22,1	12,8	12,1	507,7
pres_SUP_84ANS	0,1	5,1	11,5	125,4	21,8	6,3	0,1	0,4	NA	0,1	170,8
REG_DOM	768,7	11,1	433,8	11,7	33,8	19,9	10,9	1,6	4,2	6,2	1301,8
SEXE	60,8	30,7	121,5	96,7	162,1	NA	2,0	4,9	1,0	NA	479,7

Tableau 4.6 : Importance des variables pour l'année 2003

VARIABLE/INDICATEUR	Part auto	Part pointe am	Chaîne complexe	Mobile	1 dep	2 dep	3 dep	4 dep	5 dep	6 et plus dep	Somme
ACCES_AUTO	6127,6	28,8	5,2	517,0	15,6	25,4	22,8	2,1	6,3	4,4	6755,3
AGE	5154,4	1069,0	406,5	3725,9	3505,6	211,8	86,2	30,6	16,2	15,2	14221,3
AGE_AMP_MAX	238,0	37,4	34,4	84,0	40,6	32,5	26,4	18,5	15,2	7,8	534,9
AGE_MEAN	234,9	296,0	118,0	2821,0	930,1	150,1	91,6	32,7	17,4	10,9	4702,7
NB_FEMMES_MENAGE	237,4	13,7	23,4	22,6	27,5	10,1	18,2	4,4	2,4	6,5	366,3
NB_HOMMES_MENAGE	272,1	14,9	22,7	71,7	27,6	5,0	19,6	4,8	3,8	4,6	446,7
NB_PERSONNES_MENAGE	680,7	32,0	85,4	56,2	51,2	49,7	62,0	12,0	5,2	5,5	1039,7
pres_ENTRE12_16ANS	703,4	0,9	29,6	3,3	NA	0,2	3,2	1,4	3,2	1,5	746,8
pres_ENTRE76_84ANS	5,5	33,6	NA	609,9	148,9	34,2	0,3	NA	0,1	0,0	832,5
pres_INF_12ANS	157,2	1,9	209,0	NA	382,5	6,1	144,2	30,4	10,4	3,6	945,3
pres_SUP_84ANS	1,6	6,2	NA	188,0	43,4	8,8	0,1	NA	0,0	0,0	248,3
REG_DOM	475,8	3,2	4,7	19,0	2,1	7,5	8,5	6,1	2,7	5,6	535,2
SEXE	NA	NA	3,1	NA	211,0	4,3	9,1	4,2	0,6	1,2	233,6

Tableau 4.7 : Importance des variables pour l'année 2008

VARIABLE/INDICATEUR	Part auto	Part pointe am	Chaîne compl exe	Mobile	1 dep	2 dep	3 dep	4 dep	5 dep	6 et plus dep	Somme
ACCES_AUTO	5639,7	30,3	27,3	766,2	14,1	63,9	6,0	17,4	10,0	2,8	6577,6
AGE	4488,5	1072,5	368,5	4388,5	3516,1	290,7	78,8	44,1	40,0	13,0	14300,8
AGE_AMP_MAX	603,7	15,8	19,1	371,2	71,9	90,2	15,2	24,7	31,7	7,8	1251,3
AGE_MEAN	365,4	605,0	124,5	3672,6	1104,6	250,5	120,9	63,0	39,6	16,7	6362,8
NB_FEMMES_MENAGE	529,7	6,6	17,0	79,6	35,5	44,4	12,4	9,3	9,6	3,5	747,5
NB_HOMMES_MENAGE	549,1	8,1	18,1	163,0	34,4	22,5	15,2	12,0	8,4	5,6	836,5
NB_PERSONNES_MENAGE	1155,5	13,4	56,6	311,5	110,5	81,7	43,9	17,3	10,9	0,7	1802,2
pres_ENTRE12_16ANS	541,7	3,5	35,0	7,9	NA	18,9	5,2	6,4	3,1	0,8	622,4
pres_ENTRE76_84ANS	26,6	77,9	NA	882,4	225,6	43,2	NA	0,6	0,2	0,0	1256,4
pres_INF_12ANS	480,3	5,0	166,4	8,2	339,7	37,7	157,6	23,7	3,9	1,3	1223,9
pres_SUP_84ANS	6,7	18,1	NA	355,5	74,1	13,9	NA	0,2	0,1	0,1	468,7
REG_DOM	978,2	4,5	7,5	209,3	36,7	81,6	4,1	16,8	9,0	4,7	1352,5
SEXE	77,0	2,7	6,0	119,7	178,1	12,1	4,7	2,8	2,1	2,4	407,5

Tableau 4.8 : Importance des variables pour l'année 2013

VARIABLE/INDICATEUR	Part auto	Part pointe am	Chaîne compl exe	Mobile	1 dep	2 dep	3 dep	4 dep	5 dep	6 et plus dep	Somme
ACCES_AUTO	6668,4	2,1	22,1	528,1	25,4	5,7	0,1	2,7	1,7	0,4	7256,6
AGE	4385,9	1443,5	866,8	5146,4	4719,9	217,0	321,7	45,7	24,6	7,6	17179,1
AGE_AMP_MAX	721,4	57,6	50,2	801,8	124,7	32,4	2,4	6,3	3,9	1,2	1801,9
AGE_MEAN	559,8	832,9	335,0	4083,1	1539,1	131,7	84,4	51,7	10,6	11,2	7639,4
NB_FEMMES_MENAGE	130,8	3,2	62,2	35,5	74,3	1,8	17,5	7,3	1,6	2,5	336,7
NB_HOMMES_MENAGE	400,4	4,2	20,4	86,0	65,0	0,5	18,3	7,2	1,8	1,9	605,7
NB_PERSONNES_MENAGE	1350,2	45,6	163,1	508,2	243,7	21,5	56,2	14,7	4,0	1,3	2408,5
pres_ENTRE12_16ANS	412,6	6,8	118,0	2,7	NA	1,3	27,0	1,4	2,0	0,4	572,3
pres_ENTRE76_84ANS	NA	113,9	0,3	1157,4	257,8	24,9	NA	NA	0,0	NA	1554,3
pres_INF_12ANS	117,8	54,0	411,1	NA	817,6	1,8	126,3	58,4	9,1	3,9	1599,9
pres_SUP_84ANS	NA	28,2	0,2	461,0	91,0	9,4	NA	NA	0,0	NA	589,9
REG_DOM	932,8	6,8	5,7	65,3	5,5	8,1	2,4	3,6	1,3	1,7	1033,2
SEXE	NA	NA	26,3	NA	NA	0,2	NA	4,2	1,8	2,3	34,8

Tout d'abord, il est possible de distinguer, grâce à ces tableaux, deux sortes de variables parmi les plus importantes. Celles, comme l'accès à l'automobile, qui sont très importantes pour un indicateur (en l'occurrence la part de déplacements en automobile) et celles qui sont modérément

importantes pour plusieurs voire tous les indicateurs. Dans cette étude, il a été préféré de retenir cette dernière catégorie de variables. Ainsi, l'âge de la personne ainsi que l'âge moyen du ménage ressortent clairement comme les variables les plus importantes dans les différents arbres de décision. L'amplitude maximale des âges et le nombre de personnes dans le ménage sont aussi importants pour toutes les années d'enquête. Ensuite, la présence de moins de 12 ans et de 76 à 84 ans apparaissent comme variables importantes à partir de 2003 et ce, de manière de plus en plus prononcée jusqu'en 2013. La présence de 12 à 16 ans est importante pour l'année 2003, mais pas pour les autres années, de même que le nombre d'hommes dans le ménage en 1998. Enfin, la région du domicile est une variable plutôt importante pour l'ensemble des années. Le genre de la personne, alors qu'il est souvent cité dans la littérature comme ayant une influence sur les comportements de déplacement, ne ressort pas comme une variable importante. Ainsi, après analyse de l'importance des variables pour les quatre années d'enquête, les caractéristiques du ménage retenues comme pertinentes pour la création de typologies décrivant différents comportements de déplacements sont les suivantes :

- Âge moyen du ménage;
- Amplitude maximum des âges dans le ménage;
- Nombre de personnes dans le ménage;
- Présence d'enfants de moins de 12 ans dans le ménage;
- Présence de personnes ayant entre 76 et 84 ans dans le ménage;

4.3 Preuve que les variables sélectionnées ont bien une influence sur les déplacements des personnes

Avant de créer la typologie de ménages se basant sur la sélection de variables précédente, il a été souhaité de vérifier que les comportements de déplacement de deux personnes, dont les caractéristiques personnelles sont similaires, mais qui appartiennent à des ménages différents, sont bien différents. La Figure 4.17 positionne cette étape dans le schéma méthodologique.

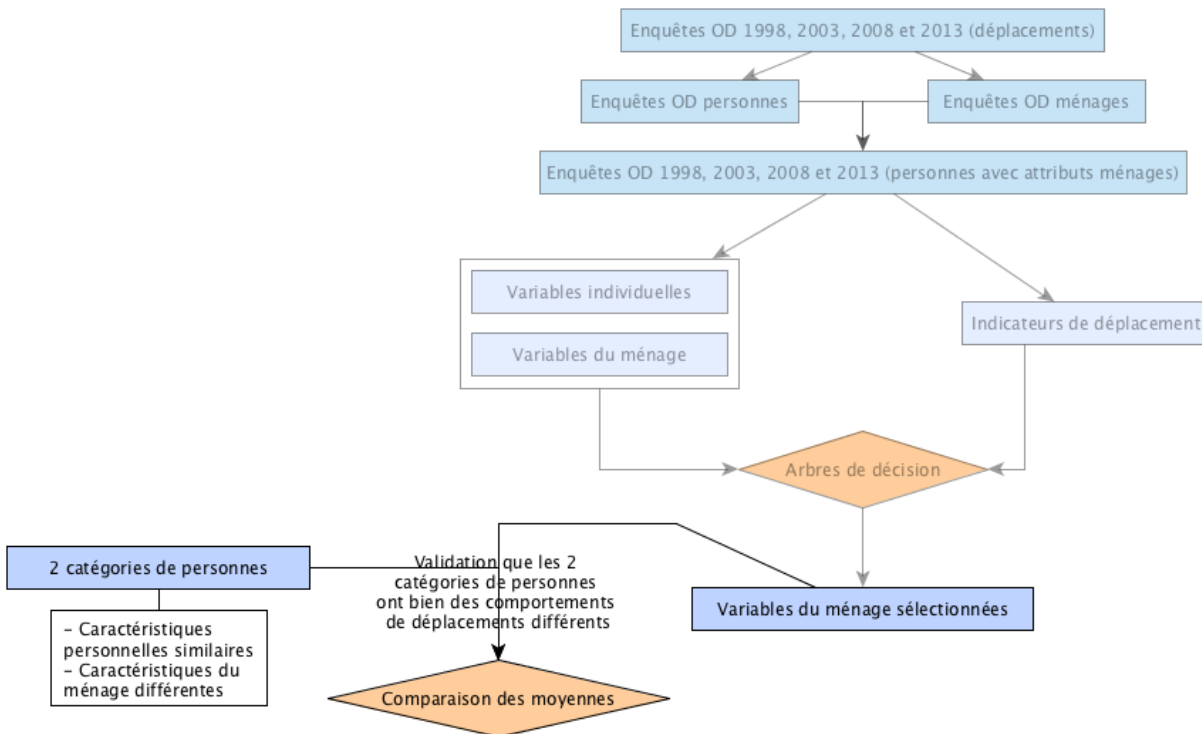


Figure 4.17 : Position de la section suivante dans le schéma méthodologique

4.3.1 Combinaisons de types de ménages

En prenant en compte toutes les valeurs que chacune des variables sélectionnées dans la section 4.2.2 peut prendre, les différents types de personnes observés, simplement avec les variables sélectionnées pour cette étude, sont très nombreux. En effet, celles-ci sont détaillées ci-dessous :

- Nombre de personnes dans le ménage : de 1 à 14, nous agrégerons ici les ménages de 6 personnes et plus, il y a donc six catégories de ménage;
- Âge moyen du ménage (de 0 à 99 ans par intervalle de 5 ans), les intervalles 0-4 ans et 5-9 ans ne seront pas considérés ici : 18 intervalles;
- Amplitude maximum des âges dans le ménage (de 0 à 100 ans par intervalle de 5 ans), les intervalles 90-94 ans et 95-99 ans ne seront pas considérés ici : 18 intervalles;
- Présence de moins de 12 ans : oui ou non;
- Présence de 76 à 84 ans : oui ou non.

Les catégories possibles de ménages sont donc au nombre de $6 \times 18 \times 18 \times 2 \times 2 = 7776$. Les catégories de ménages réellement présentes (ici, l'étude est faite sur l'enquête OD 2013) sont au nombre de

998. Parmi celles-ci, seulement 137 représentent au moins 1% de la population et 35 représentent au moins 5% de la population.

Cependant, chaque personne a un âge entre 0 et 99 ans et est soit du genre masculin soit du genre féminin, ce qui multiplie encore le nombre de personnes qui pourraient être comparées ici. Il est évidemment impossible de comparer tous les types de personnes de cette enquête. Deux catégories de personnes ont été choisies, celles-ci possèdent des caractéristiques individuelles similaires, mais appartiennent à des ménages différents. Ceci est fait dans le but de montrer que la composition du ménage a bien une influence sur les déplacements de la personne.

4.3.2 Comparaison des déplacements de deux catégories de personnes

Les hommes âgés de 25 à 34 ans ont été choisis pour cette comparaison; deux catégories de ménages vont également être choisies. Les catégories sélectionnées doivent posséder un nombre d'observations assez élevé (au moins supérieur à 30). Ainsi, parmi les hommes de 25 à 34 ans, les quinze catégories de ménages les plus représentées sont montrées dans la Figure 4.18. Les caractéristiques de chaque catégorie sont détaillées dans le Tableau 4.9.

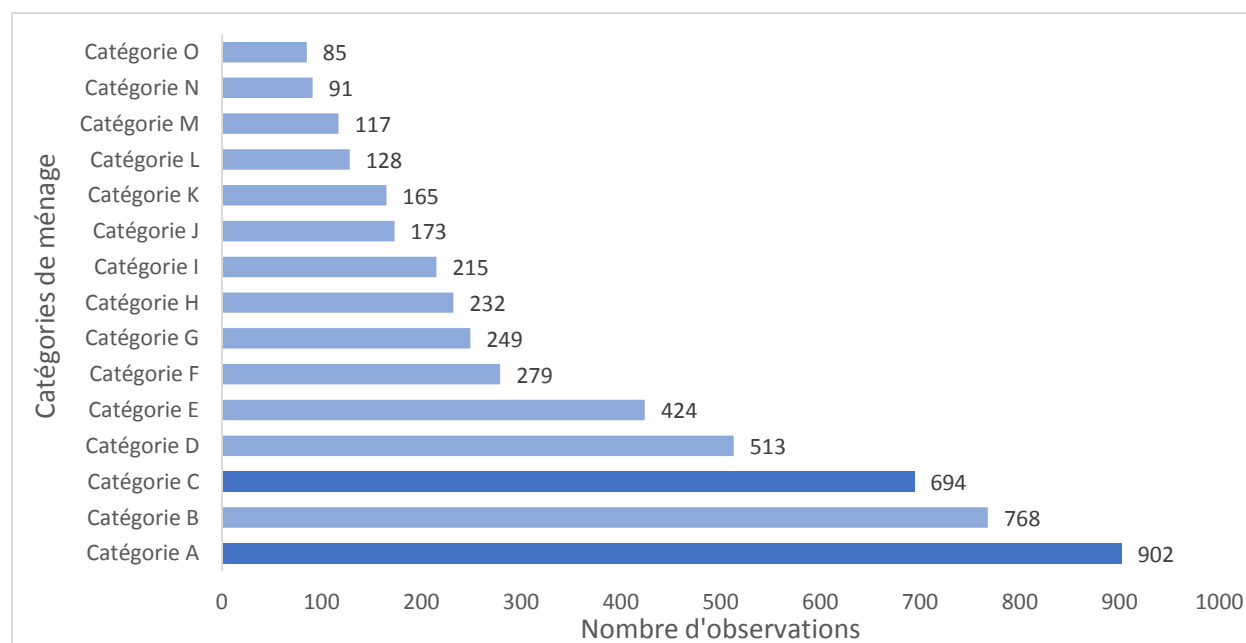


Figure 4.18 : 15 principales catégories de ménage observées pour les hommes de 25 à 34 ans

Tableau 4.9 : Propriétés des catégories de ménage les plus fréquentes

Catégorie A	Âge moyen [25,30); Amp max [0,5); pas d'enfant, pas de personnes âgées, 2 personnes
Catégorie B	Âge moyen [30,35); Amp max [0,5); pas d'enfant, pas de personnes âgées, 2 personnes
Catégorie C	Âge moyen [15,20); Amp max [30,35); présence d'enfant, pas de personnes âgées, 4 personnes
Catégorie D	Âge moyen [20,25); Amp max [30,35); présence d'enfant, pas de personnes âgées, 3 personnes
Catégorie E	Âge moyen [30,35); Amp max [0,5); pas d'enfant, pas de personnes âgées, 1 personne
Catégorie F	Âge moyen [15,20); Amp max [25,30); présence d'enfant, pas de personnes âgées, 4 personnes
Catégorie G	Âge moyen [25,30); Amp max [0,5); pas d'enfant, pas de personnes âgées, 1 personne
Catégorie H	Âge moyen [20,25); Amp max [25,30); présence d'enfant, pas de personnes âgées, 3 personnes
Catégorie I	Âge moyen [15,20); Amp max [25,30); présence d'enfant, pas de personnes âgées, 3 personnes
Catégorie J	Âge moyen [45,50); Amp max [30,35); pas d'enfant, pas de personnes âgées, 3 personnes
Catégorie K	Âge moyen [25,30); Amp max [5,10); pas d'enfant, pas de personnes âgées, 2 personnes
Catégorie L	Âge moyen [30,35); Amp max [5,10); pas d'enfant, pas de personnes âgées, 2 personnes
Catégorie M	Âge moyen [15,20); Amp max [30,35); présence d'enfant, pas de personnes âgées, 5 personnes
Catégorie N	Âge moyen [40,45); Amp max [25,30); pas d'enfant, pas de personnes âgées, 3 personnes
Catégorie O	Âge moyen [40,45); Amp max [30,35); pas d'enfant, pas de personnes âgées, 4 personnes

Les deux catégories sélectionnées pour la comparaison sont en bleu foncé sur le graphique, leurs caractéristiques sont décrites dans le Tableau 4.10. Ainsi la catégorie 1 correspond plutôt à un ménage de type colocation, tandis que la catégorie 2, elle, ressemble davantage à une jeune famille.

Tableau 4.10 : Caractéristiques du ménage des deux catégories de personnes

	Catégorie 1 (902 observations pour l'enquête 2013)	Catégorie 2 (694 observations pour l'enquête 2013)
Âge moyen du ménage	25-30 ans	15-20 ans
Amplitude maximum des âges dans le ménage	0-5 ans	30-35 ans
Nombre de personnes dans le ménage	2	4
Présence de moins de 12 ans	Non	Oui
Présence de 76-84 ans	Non	Non

Une fois ces observations sélectionnées, une analyse des comportements de déplacements sur ce nouvel ensemble de données est effectuée.

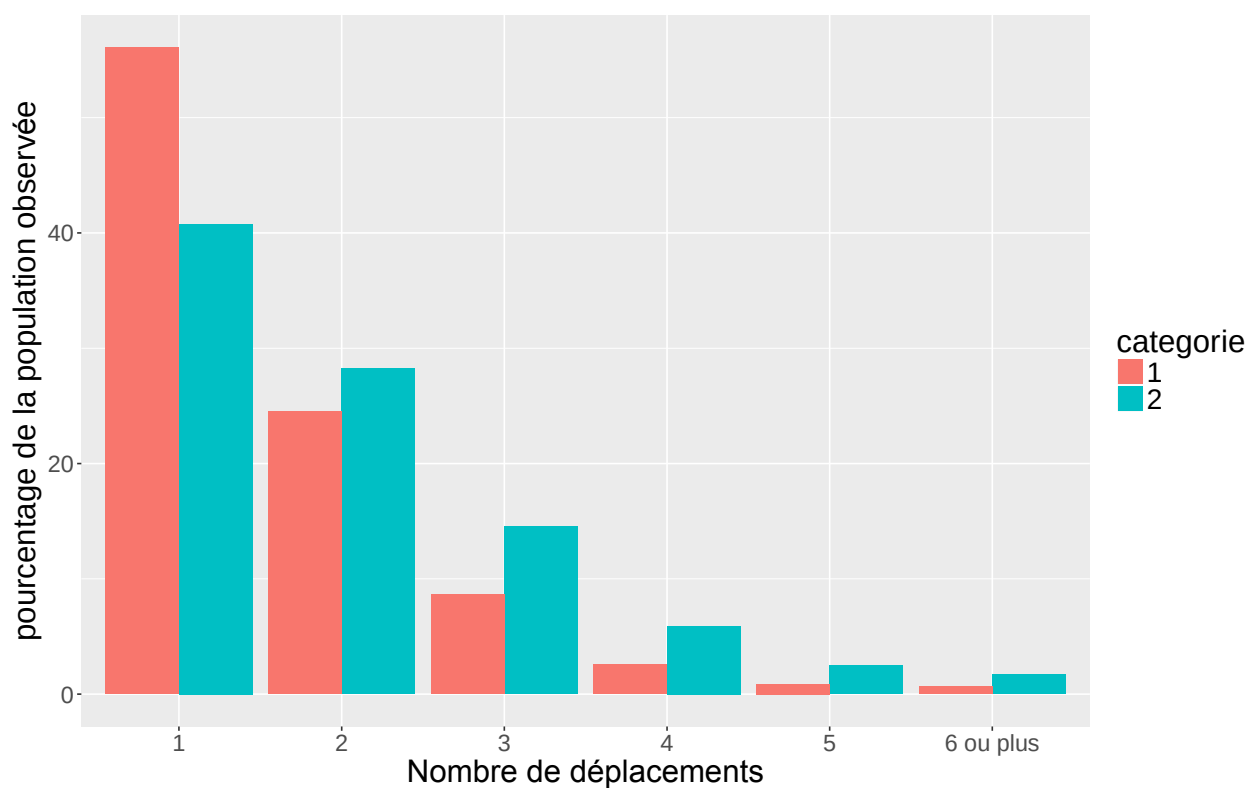


Figure 4.19 : Nombre de déplacements par personne du groupe d'âge sélectionné suivant la catégorie de ménage (enquête 2013)

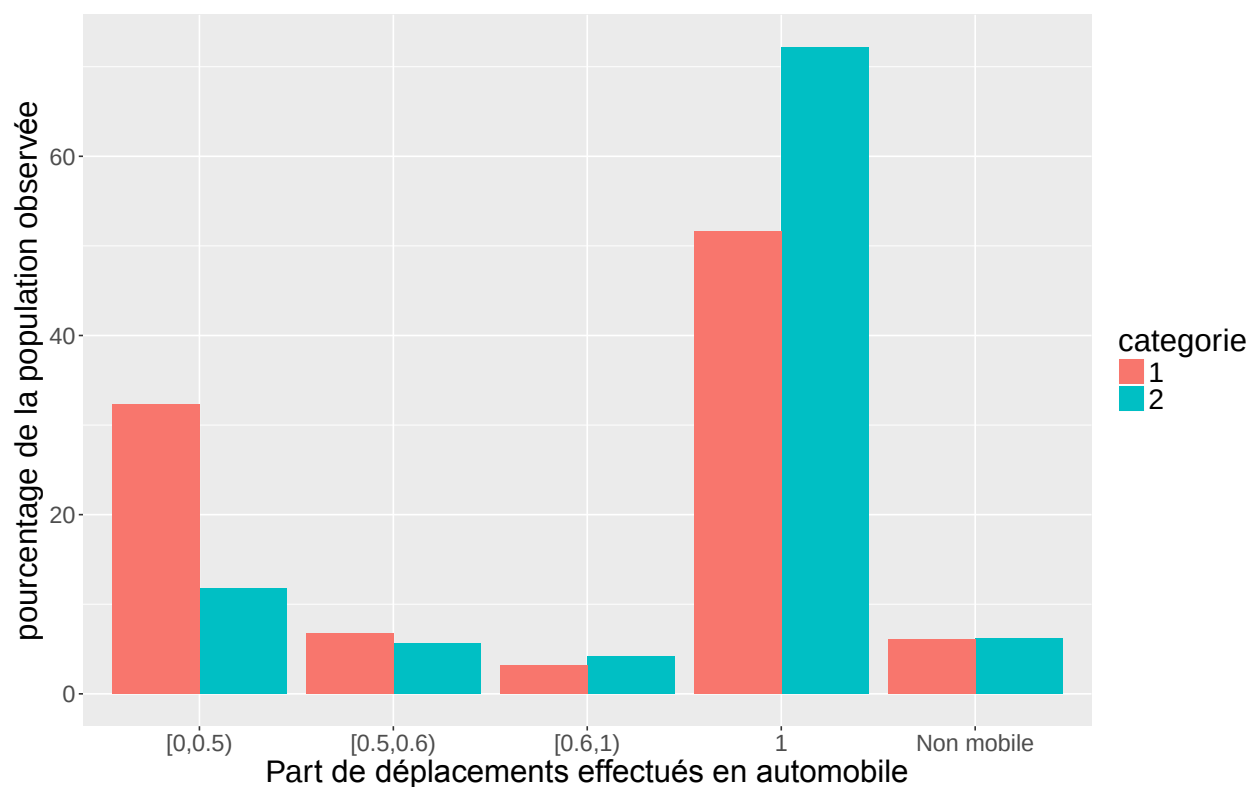


Figure 4.20 : Part de déplacements en auto par personne suivant la catégorie de ménage (enquête 2013)

De nettes différences de comportements sont effectivement observées. Celles sur le nombre de déplacements par personne apparaissent parmi les plus significatives (Figure 4.19). Les hommes de la catégorie 1, à savoir un ménage type colocation de personnes entre 20 et 35 ans, sont plus nombreux à faire peu de déplacements, tandis que pour la catégorie 2 dont le ménage s'apparente plus à un couple avec des enfants, les hommes sont plus nombreux à faire davantage de déplacements dans une journée.

Des différences sont également observées pour la part de déplacements en auto (Figure 4.20). Les hommes de la catégorie 2 font majoritairement leurs déplacements en auto alors que le contraire est observé pour les hommes de la catégorie 1.

L'étude de l'ensemble des indicateurs révèle donc certaines différences de comportement de déplacements pour les deux catégories de personnes analysées ici. Ceci est de bon augure pour la prochaine étape du projet qui consiste à élaborer la typologie de ménages.

CHAPITRE 5 TYPOLOGIES

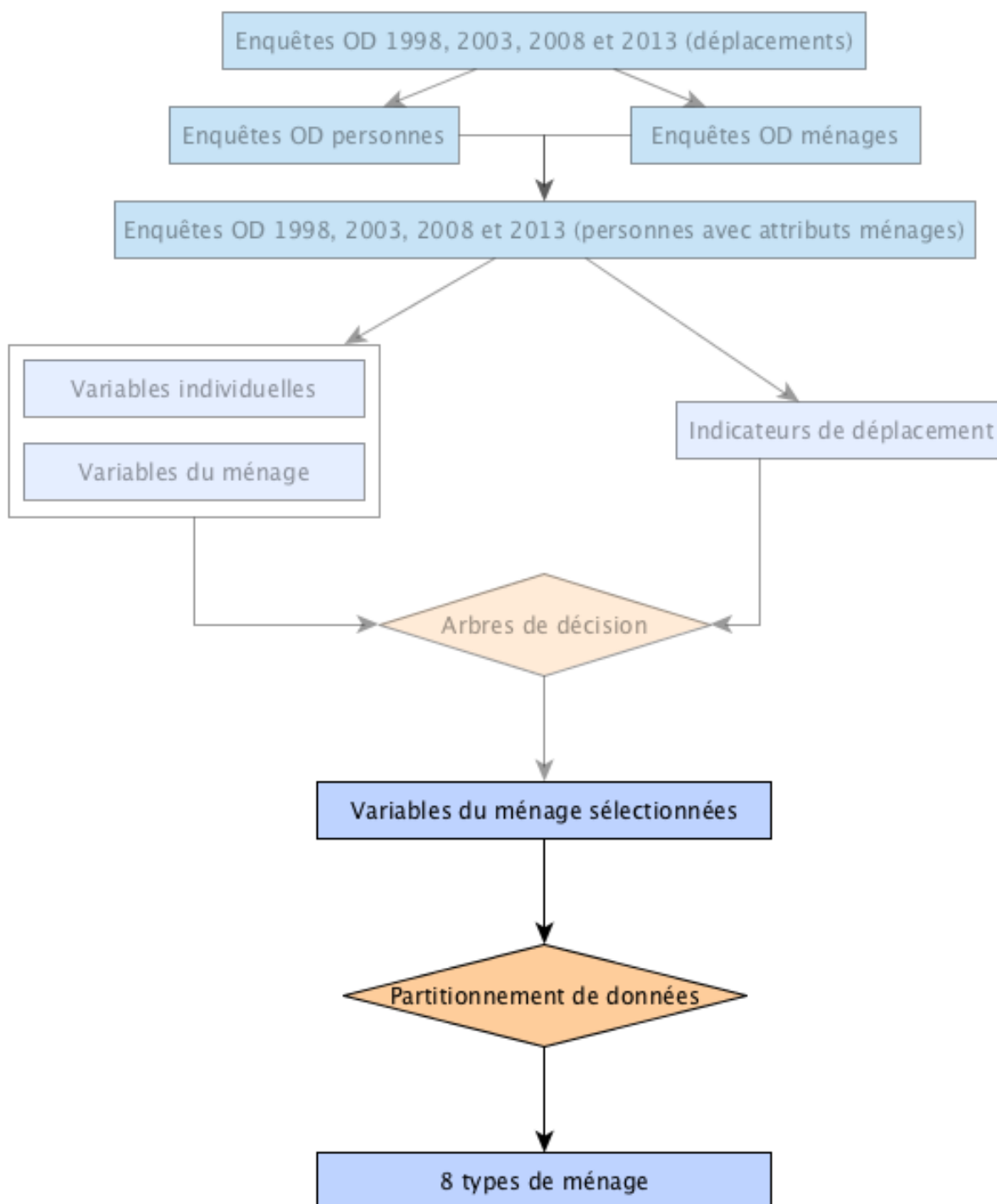


Figure 5.1 : Position de la section suivante dans le schéma méthodologique

Une fois la sélection des variables du ménage influentes pour les comportements de déplacement individuels faite, il faut créer la typologie de ménage. Celle-ci permettra de prendre en compte les caractéristiques du foyer d'un individu dans la prévision de la demande en transport. Cette étape est placée dans la méthodologie globale dans la Figure 5.1. Elle sera réalisée grâce à un algorithme de fouille de données : l'algorithme des k-moyennes, présenté dans la section 3.5. Cet algorithme a été appliqué sur les enquêtes OD de chaque année (1998, 2003, 2008 et 2013). Puis, pour une question de transférabilité temporelle, elle a également été appliquée à l'ensemble des données de toutes les enquêtes rassemblées dans un même fichier. Pour vérifier la pertinence de la typologie retenue, les comportements de déplacement des individus appartenant à différents types de ménage sont comparés.

5.1 Typologies de ménages année par année

L'algorithme des k-moyennes est premièrement appliqué aux bases de données de chaque année d'enquête individuellement. Les variables « entrantes » de l'algorithme correspondent aux caractéristiques du ménage retenues dans la section précédente. L'algorithme des k-moyennes ne peut pas être effectué sur des variables catégoriques. Ainsi, les variables « présence d'enfants de moins de 12 ans » et « présence de personnes âgées de 76 à 84 ans » seront remplacées par le nombre de ces catégories de personnes dans le ménage. Pour rappel, les variables entrantes utilisées sont :

- Le nombre de personnes dans le ménage,
- L'âge moyen du ménage,
- L'amplitude maximum des âges dans le ménage,
- Le nombre de personnes de moins de 12 ans,
- Le nombre de personnes ayant entre 76 et 84 ans.

Les variables « sortantes » de l'algorithme quant à elles correspondent à huit types de ménages dont les caractéristiques varient d'un type à un autre. Le nombre huit a été déterminé par essais-erreur afin d'obtenir un nombre de groupes suffisant et en prenant garde de ne pas avoir deux types de ménages trop similaires. Les quatre sections suivantes présentent les typologies obtenues pour chaque année.

5.1.1 Enquête 1998

Les résultats du partitionnement de données pour l'année 1998 sont décrits dans le Tableau 5.1 (taille des groupes, variance à l'intérieur des groupes, part de la variance expliquée par la variance entre les groupes et les moyennes de chaque variable).

Tableau 5.1 : Résultats du partitionnement de données pour l'enquête OD 1998

Types de ménages	Jeunes couples, colocs ou célibataires	Cinquantenaires	Retraités	Adultes et personnes âgées	Familles traditionnelles	Très jeunes familles	Jeunes familles	Familles multigénérationnelles
Taille	19542	18782	16298	9737	29424	33117	22772	4372
Variance à l'intérieur des groupes	902386,3	912122,1	997100,6	961607,6	1104715,5	978975,0	764646,1	624842,4
Variance inter/variance totale	0,913							
Nombre de personnes	1,9	1,7	1,7	3,0	3,5	4,0	4,4	4,8
Âge moyen	29,6	51,9	70,7	51,1	34,1	20,5	25,4	38,9
Amplitude max des âges	2,6	3,1	2,9	34,1	30,5	29,4	38,2	60,0
Nombre de moins de 12 ans	0,0	0,0	0,0	0,0	0,0	1,6	1,3	0,8
Nombre de 76-84 ans	0,0	0,0	0,2	0,1	0,0	0,0	0,0	0,2

Chaque type de ménage a été nommé suivant ses caractéristiques pour rendre l'interprétation plus facile. Par exemple, le premier type de ménage possède un nombre de personnes moyen de 1,9, un âge moyen autour de 30 ans et une faible amplitude maximale des âges. Aucun enfant de moins de 12 ans ou personne de 76 à 84 ans ne sont recensés dans ce groupe. Ce dernier correspond donc à des ménages de petite taille où tous les habitants sont plutôt jeunes et de la même génération; il s'apparente donc bien à des ménages de type « colocation, couples ou célibataires ». Le troisième type de ménage, lui, possède une moyenne d'âge élevée, peu de personnes et une amplitude maximale des âges faible. Il a été nommé « Retraités ». Le type 8 possède une amplitude maximale des âges très élevée, un âge moyen intermédiaire et est

composé de beaucoup de personnes. La présence de moins de 12 ans et de 76-84 ans est généralement observée. Ce sont des ménages atypiques; la taille du groupe (nombre de personnes appartenant au groupe) est d'ailleurs plus faible que les autres. On y retrouve souvent à la fois des enfants, des adultes et des personnes âgées, c'est pourquoi il a été nommé « familles multigénérationnelles ». Ici, 91% de la variance est expliquée par les différences entre les groupes, ce qui correspond avec la volonté de maximiser les écarts entre les groupes et minimiser ceux à l'intérieur des groupes.

La Figure 5.2 représente l'appartenance aux différents types de ménages suivant l'âge et le genre. Il apparaît, pour certaines catégories d'âge, que l'appartenance à un type de ménage peut facilement être prédite puisque seulement un ou deux types sont observés pour la tranche d'âge en question. Au contraire, pour d'autres catégories d'âges, trois ou quatre types de ménages prédominent, par conséquent, l'affectation à un type de ménage n'apparaît pas de manière évidente.

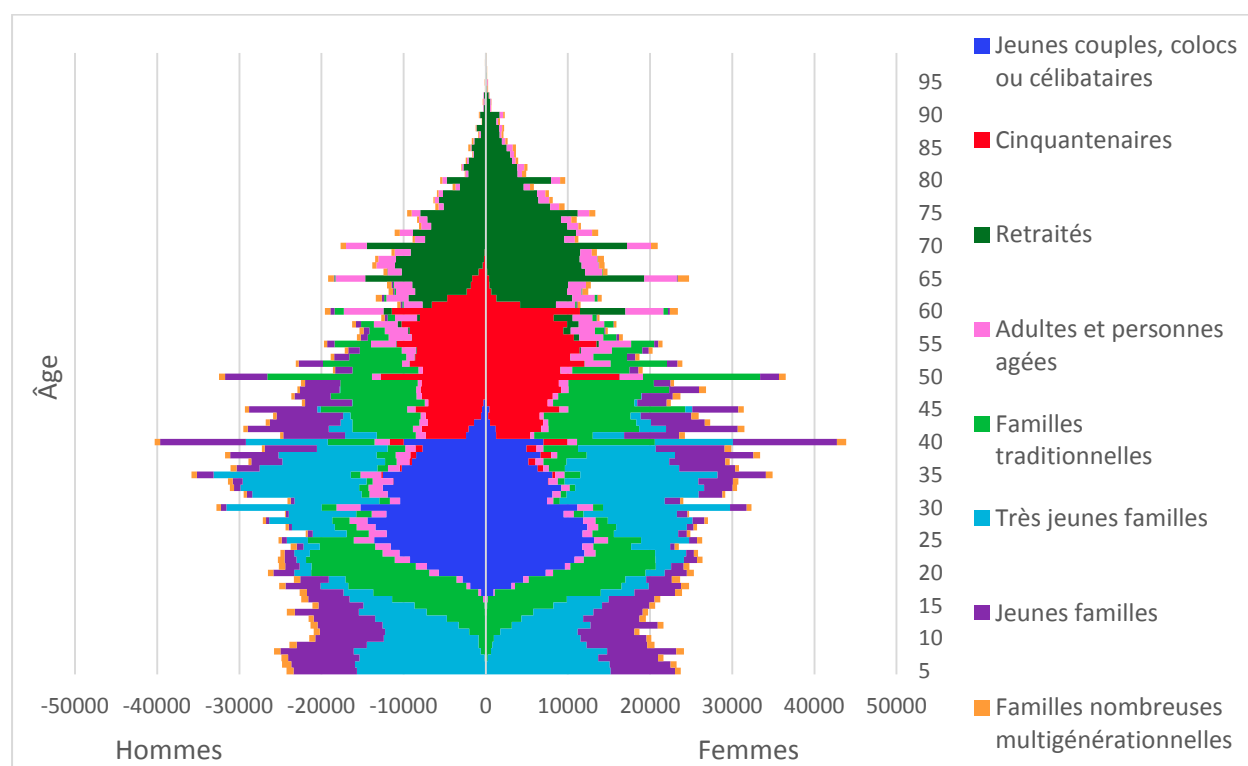


Figure 5.2 : Appartenance au type de ménage suivant l'âge et le genre (enquête 1998)

Il est à noter que sur les graphiques de l'appartenance d'une personne à un type de ménage suivant son âge et son genre, des « pics » sont observés aux dizaines (40 ans, 50 ans, 60 ans, ...).

Ceci est dû au fait que les personnes interviewées ne donnent pas nécessairement leur âge exact, mais le groupe d'âge auquel ils appartiennent. Ces pics sont moins observés pour les dizaines en bas âges, car les personnes plus jeunes ont plus tendance à donner leur âge exact.

5.1.2 Enquête 2003

Pour l'année 2003, les résultats du partitionnement des données sont similaires à ceux de 1998. Ils sont présentés dans le Tableau 5.2. Il en est de même pour l'appartenance à un type de ménage suivant l'âge et le genre, présentée en Figure 5.3.

Tableau 5.2 : Résultats du partitionnement de données pour l'enquête OD 2003

Types de ménages	Jeunes couples, colocs ou célibataires	Cinquanténaires	Retraités	Adultes et personnes âgées	Familles traditionnelles	Très jeunes familles	Jeunes familles	Familles multigénérationnelles
Taille	22187	25175	18001	7441	28627	33292	25083	4540
Variance à l'intérieur des groupes	1029683,5	1217559,8	1068208,8	788270,5	1183964,0	1053765,9	844451,7	638894,8
Variance inter/variance totale	0,916							
Nombre de personnes	1,9	1,7	1,7	2,9	3,4	4,0	4,4	4,7
Âge moyen	29,6	53,0	71,6	54,3	35,8	21,4	26,1	39,2
Amplitude max des âges	2,7	3,0	2,9	32,7	31,3	29,8	38,6	58,4
Nombre de moins de 12 ans	0,0	0,0	0,0	0,0	0,0	1,4	1,2	0,8
Nombre de 76-84 ans	0,0	0,0	0,3	0,2	0,0	0,0	0,0	0,2

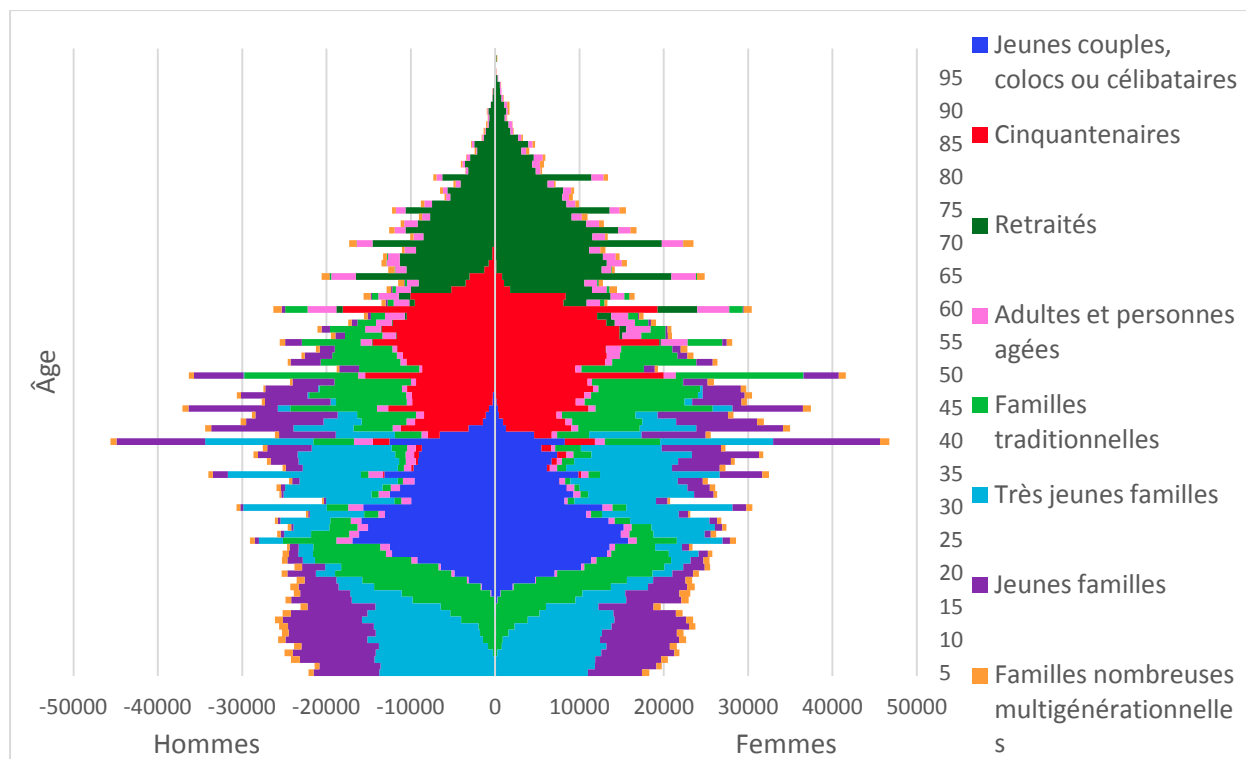


Figure 5.3 : Appartenance au type de ménage suivant l'âge et le genre (enquête 2003)

5.1.3 Enquête 2008

Pour 2008, les résultats du partitionnement de données sont encore une fois semblables. L'appartenance à un type de ménage suivant l'âge et le genre pour l'enquête 2008 est présentée en Figure 5.4.

Tableau 5.3 : Résultats du partitionnement de données pour l'enquête OD 2008

Types de ménages	Jeunes couples, colocs ou célibataires	Cinquantennaires	Retraités	Adultes et personnes âgées	Familles traditionnelles	Très jeunes familles	Jeunes familles	Familles multigénérationnelles
Taille	14092	26290	20147	5742	26939	28886	22456	4406
Variance à l'intérieur des groupes	726749,1	1144014,7	1152647,4	613499,7	1238560,1	1024314,8	784899,0	646457,5
Variance inter/variance totale	0,921							
Nombre de personnes	1,8	1,7	1,6	2,8	3,3	4,0	4,3	4,4
Âge moyen	32,0	55,4	73,3	60,2	37,5	22,1	27,0	41,0
Amplitude max des âges	2,8	2,7	2,4	30,4	32,5	30,6	39,2	58,2
Nombre de moins de 12 ans	0,0	0,0	0,0	0,0	0,0	1,3	1,1	0,6
Nombre de 76-84 ans	0,0	0,0	0,4	0,3	0,0	0,0	0,0	0,2

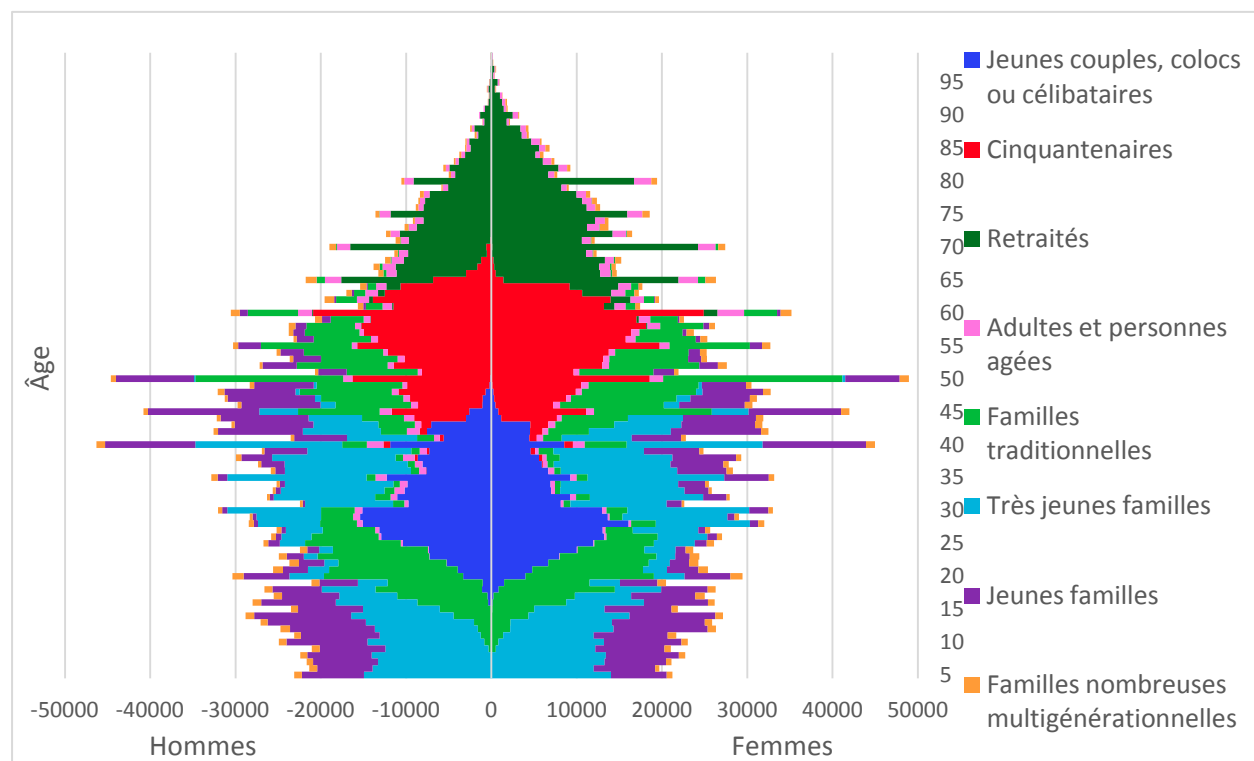


Figure 5.4 : Appartenance au type de ménage suivant l'âge et le genre (enquête 2008)

5.1.4 Enquête 2013

En 2013, le partitionnement de données mène à de légères nuances par rapport aux années précédentes. Quatre types de ménages (au lieu de trois) se démarquent pour les types de ménages ayant des amplitudes maximales des âges très faibles. Les « cinquantenaires » et « retraités » des années précédentes se décomposent maintenant en « cinquantenaires », « soixantenaires » et « retraités ». Au contraire, les « jeunes familles » et « très jeunes familles » sont, en 2013, regroupées en un seul type de ménage. Ceci est probablement dû au vieillissement de la population. En effet, l'augmentation du nombre de personnes dans les types de ménages où la moyenne d'âge est plus élevée pousse l'algorithme à diviser ces derniers en plusieurs groupes.

Tableau 5.4 : Résultats du partitionnement de données pour l'enquête OD 2013

Types de ménages	Jeunes couples, colocs ou célibataires	Cinquantenaires	Soixantenaires	Retraités	Adultes et personnes âgées	Familles traditionnelles	Jeunes familles	Familles multigénérationnelles
Taille	11809	23224	25559	12975	7942	39854	50246	7299
Variance à l'intérieur des groupes	589053,3	741641,2	632034,3	510571,0	916985,5	1986029,5	2295159,5	1116247,9
Variance inter/variance totale	0,926							
Nombre de personnes	1,9	1,8	1,8	1,6	2,8	3,5	4,2	4,5
Âge moyen	32,2	54,0	66,1	79,3	59,3	36,8	23,0	39,4
Amplitude max des âges	2,8	2,7	2,9	2,5	30,6	34,4	34,4	57,1
Nombre de moins de 12 ans	0,0	0,0	0,0	0,0	0,0	0,1	1,5	0,8
Nombre de 76-84 ans	0,0	0,0	0,0	0,8	0,3	0,0	0,0	0,2

Pour ce qui est de la répartition des personnes dans les types de ménages suivant leur âge et leur genre (Figure 5.5), elle est similaire aux années précédentes. Il est toujours possible d'observer de nombreuses possibilités de types de ménages pour certaines catégories d'âge et moins pour d'autres. Cependant, l'âge à partir duquel le nombre de types de ménages auxquels une personne

est susceptible d'appartenir diminue est très légèrement plus bas pour l'enquête 2013 que pour les précédentes (autour de 55 ans contre 60 ans pour les années précédentes).

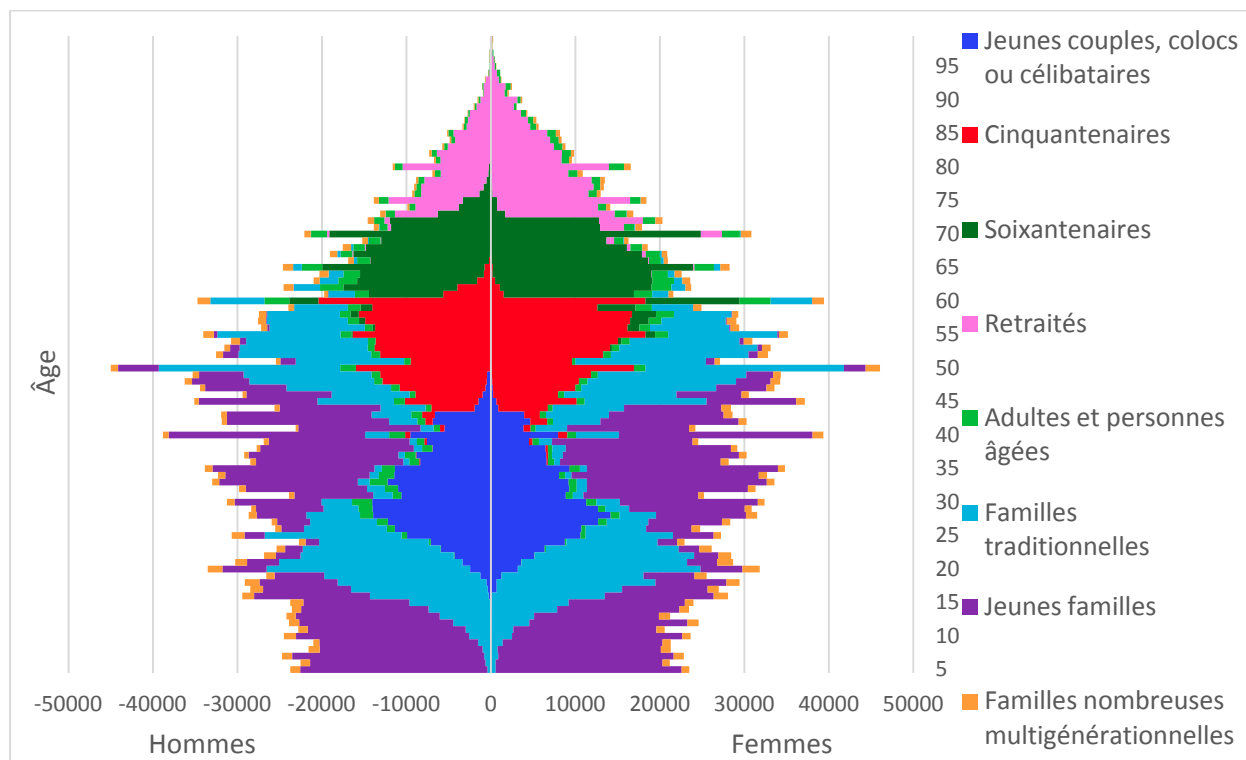


Figure 5.5 : Appartenance au type de ménage suivant l'âge et le genre (enquête 2013)

5.1.5 Transférabilité temporelle

Puisque le but de la création d'une typologie de ménages est de mieux prévoir la demande de mobilité future, il est nécessaire que cette dernière soit transférable dans le temps. Or, si pour les années 1998, 2003 et 2008, les groupes formés pour la typologie sont similaires (seules les frontières et les moyennes des groupes varient d'une année à l'autre), ce n'est pas le cas pour l'année 2013. La Figure 5.6 représente les variations des variables « âge moyen » et « amplitude maximale des âges » dans le ménage entre 1998 et 2003. Les enveloppes convexes sont utilisées pour représenter les personnes dans chaque groupe de ménage. La Figure 5.7 présente les mêmes variations, mais entre les années 2003 et 2008. Il est possible de voir que l'âge moyen de chaque groupe semble augmenter, particulièrement pour les ménages ayant une petite amplitude des âges, les frontières des groupes se décalent vers la droite. Mis à part cette différence, les groupes formés sont très similaires. Cela se voit également lorsque l'on compare les moyennes de chaque variable (Tableau 5.1, Tableau 5.2 et Tableau 5.3).

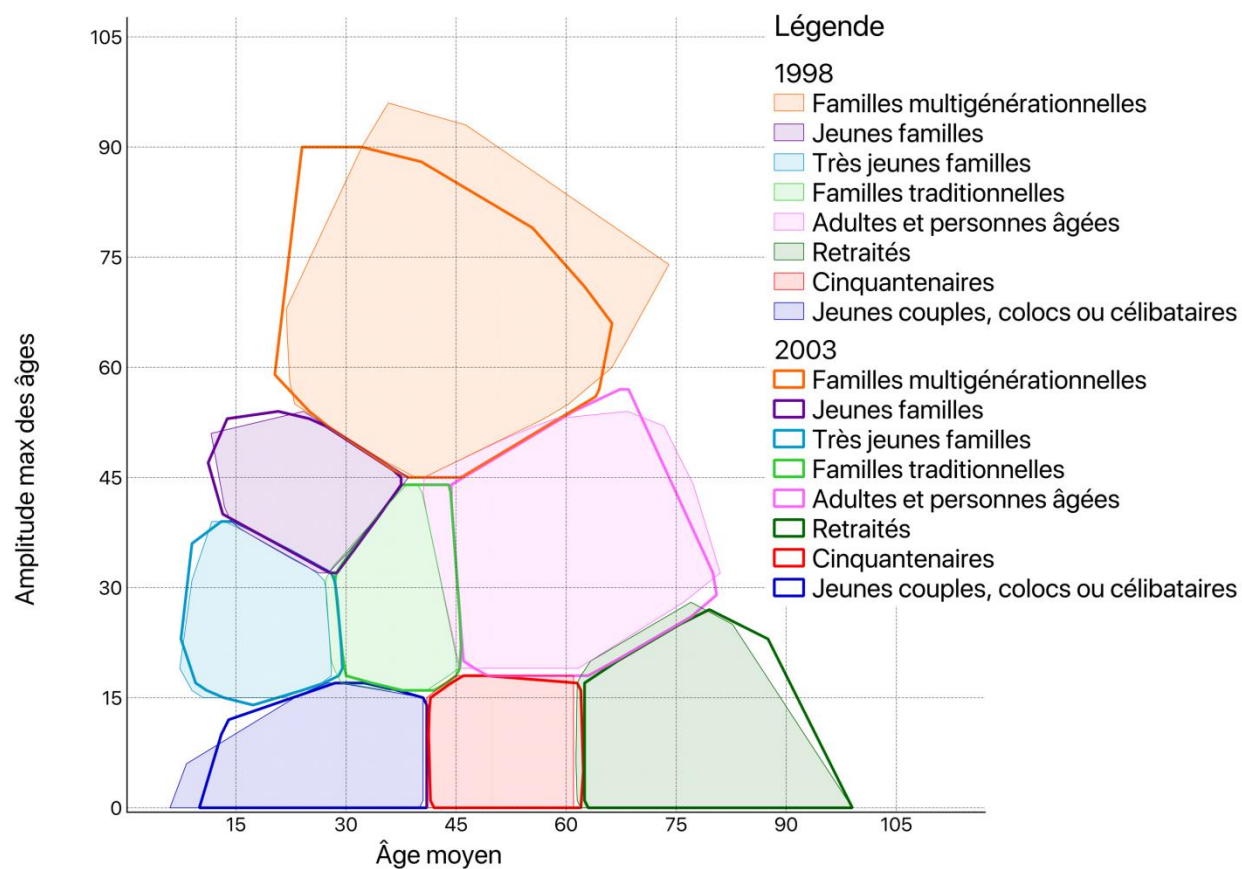


Figure 5.6 : Répartition des groupes suivant l'âge moyen et l'amplitude max des âges, enquêtes 1998 et 2003

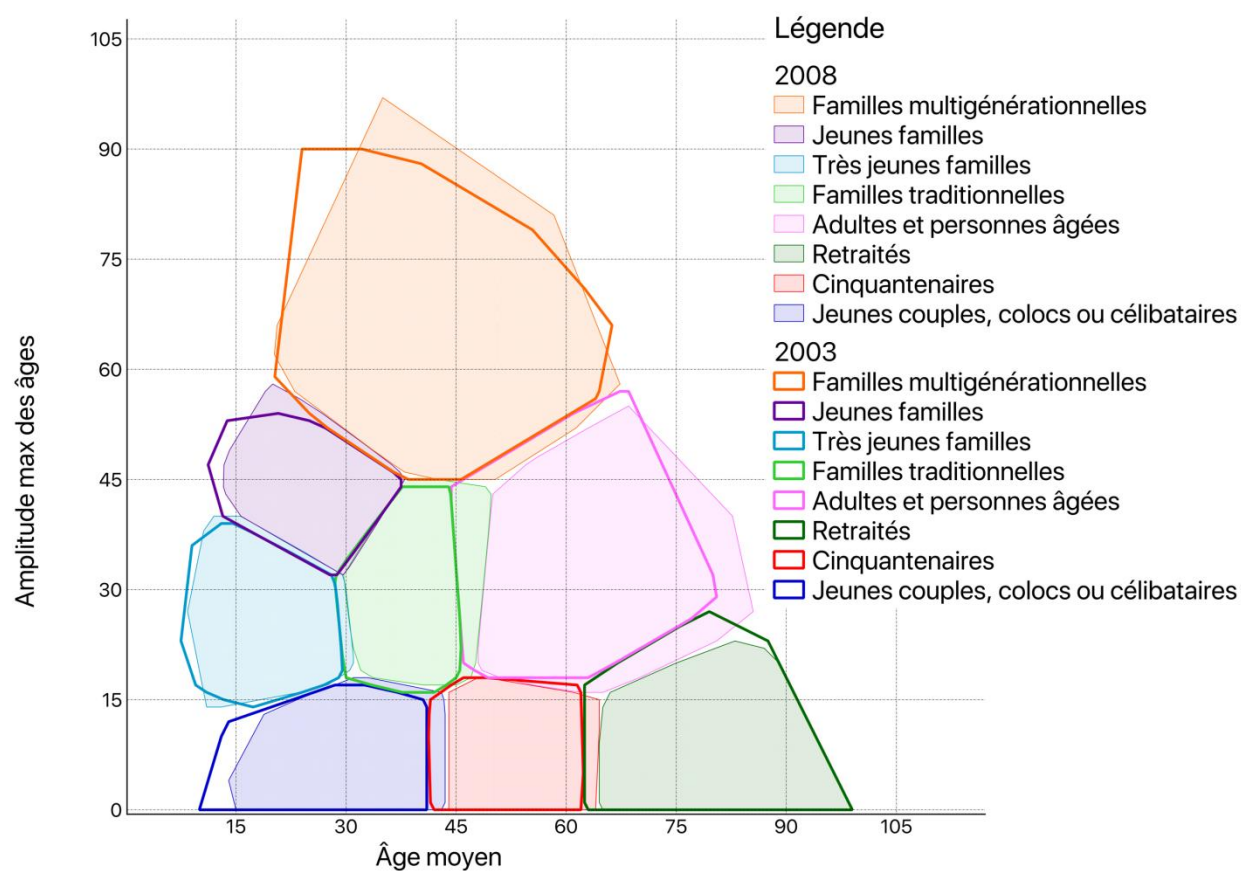


Figure 5.7 : Répartition des groupes suivant l'âge moyen et l'amplitude max des âges, enquêtes 2003 et 2008

Sur la Figure 5.8, il est possible d'observer les changements qui apparaissent dans la typologie de 2013. La séparation des types « cinquantenaires » et « retraités », ainsi que la fusion des « jeunes familles » et « très jeunes familles » y sont bien visibles. Aussi, en se référant au Tableau 5.4 et à ceux cités précédemment, il apparaît que même pour les types de ménages « constants », les moyennes des variables continuent d'évoluer sur les mêmes tendances que pour les trois premières années. Par exemple, l'âge moyen des « retraités » passe de 70,7 ans en 1998 à 71,6 ans en 2003 puis à 73,3 ans en 2008 et il atteint 79,3 ans en 2013. Le nombre de personnes moyen pour les familles multigénérationnelles lui, diminue progressivement (4,5 personnes en 2013 contre 4,8 personnes en 1998).

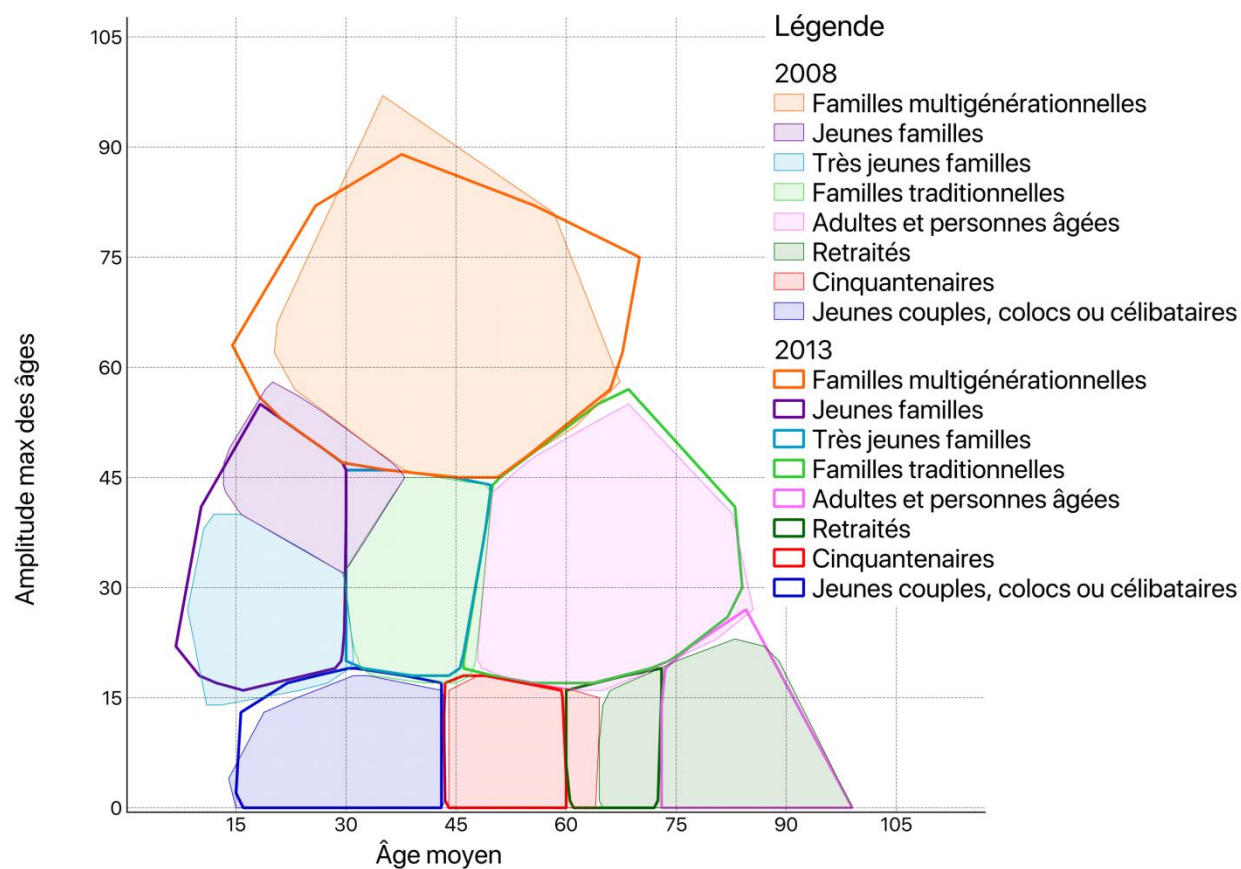


Figure 5.8 : Répartition des groupes suivant l'âge moyen et l'amplitude max des âges, enquêtes 2008 et 2013

Prévoir la création de nouveaux types de ménage ou bien la fusion de certains n'est pas une chose aisée. Ainsi, la transférabilité temporelle n'est pas validée dans cette situation. Il a donc été choisi de créer une typologie unique pour toutes les années, où l'appartenance d'une personne à un groupe dépendrait de l'année d'enquête.

5.2 Typologie retenue

Pour créer cette typologie unique, les données de toutes les enquêtes de 1998 à 2013 ont été rassemblées dans une seule et même table où la variable « année d'enquête » a été ajoutée. Ensuite, une typologie a été créée sur cette base de données.

5.2.1 Typologie sur 4 années d'enquête

Les résultats du partitionnement de données sur la base de données comprenant les quatre années d'enquête ressemblent davantage à ceux des années 1998 à 2008 qu'à ceux de 2013. Les mêmes noms de groupes que pour les plus anciennes enquêtes ont donc été donnés. Le Tableau 5.5 recense les moyennes des variables pour ce partitionnement ainsi que la taille et la variance des groupes. La Figure 5.9 représente l'amplitude maximale des âges dans le ménage en fonction de l'âge moyen de celui-ci. Les âges moyens de chaque groupe et le nombre de personnes sont très similaires dans le partitionnement de l'ensemble des données et dans ceux des années 1998 à 2008. De même, l'appartenance à un type de ménage suivant l'âge et le genre (Figure 5.10) est également semblable à celles des années 2008 et précédentes. Il est encore possible d'observer la différence dans le nombre de types de ménages auxquels les personnes sont susceptibles d'appartenir selon leur catégorie d'âge.

Tableau 5.5 : Résultats du partitionnement de données sur l'ensemble des données (1998 à 2013)

Types de ménages	Jeunes couples, colocs ou célibataires	Cinquantennaires	Retraités	Adultes et personnes âgées	Familles traditionnelles	Très jeunes familles	Jeunes familles	Familles multigénérationnelles
Taille	69251	102273	82364	26915	117727	129566	97992	20163
Variance à l'intérieur des groupes	3545938,0	4624632,0	4752147,0	2966272,0	5322576,0	4358357,0	3342749,0	2952850,0
Variance inter/variance totale	0,919							
Nombre de personnes	1,9	1,8	1,7	2,8	3,4	4,0	4,4	4,6
Âge moyen	30,8	54,8	72,6	57,5	36,9	21,6	26,3	40,1
Amplitude max des âges	2,7	2,9	2,6	31,7	32,3	30,1	38,8	58,4
Nombre de moins de 12 ans	0,0	0,0	0,0	0,0	0,0	1,4	1,2	0,7
Nombre de 76-84 ans	0,0	0,0	0,3	0,2	0,0	0,0	0,0	0,2

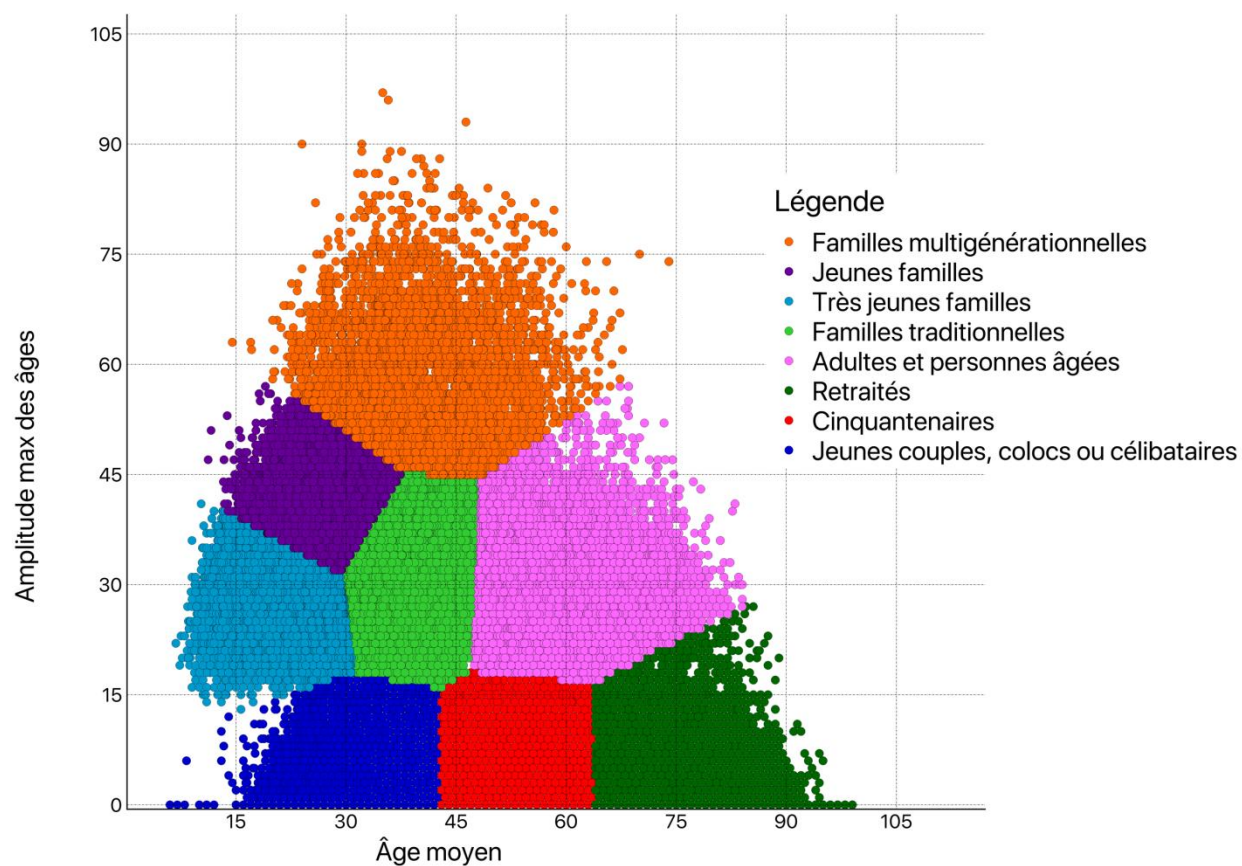


Figure 5.9 : Répartition des personnes suivant l'amplitude maximale des âges et l'âge moyen dans leur ménage (1998 à 2013)

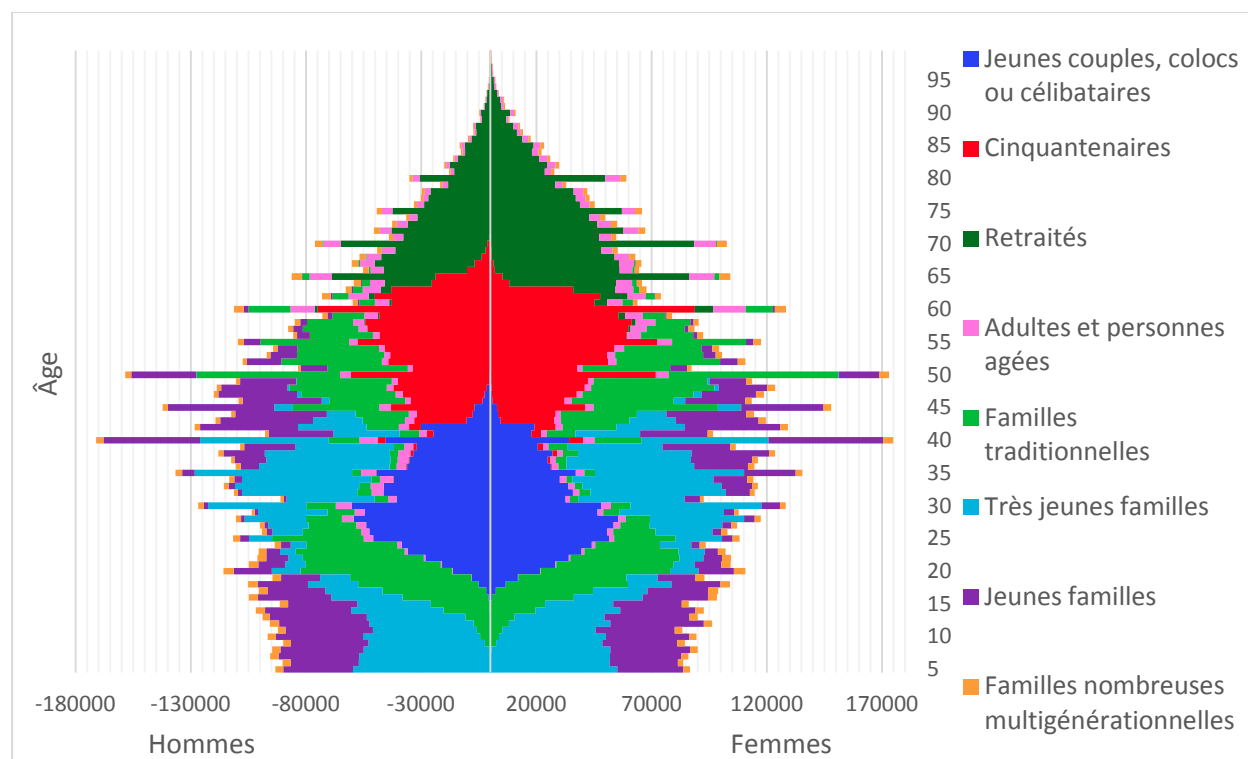


Figure 5.10 : Appartenance au type de ménage suivant l'âge et le genre (enquêtes 1998 à 2013)

Il est alors nécessaire de se demander si les changements dans la population, matérialisés par les variations dans la formation des types de ménage d'une année à l'autre, restent observables dans la typologie créée pour les quatre années d'enquête. Ainsi, la répartition des personnes dans les types de ménages est étudiée à travers le temps. La Figure 5.11 représente le pourcentage de la population appartenant à chaque type de ménage par année. Plusieurs faits en ressortent. Premièrement, il est plus facile de visualiser l'importance (en nombre de personnes) des groupes. Les « familles multigénérationnelles » et les « adultes et personnes âgées » représentent une très petite part de la population; cela est intuitif étant donné que ce sont des ménages plus atypiques. Au contraire, les types de ménages de familles représentent une plus grosse part de la population. Deuxièmement, des évolutions à travers les années sont visibles. Le pourcentage des « retraités » augmente considérablement et continuellement de 1998 à 2013. De même, la part des « jeunes familles » et « familles traditionnelles » augmente aux dépens des « très jeunes familles ». Ces deux phénomènes sont des résultantes du vieillissement de la population. Les « jeunes couples, colocs ou célibataires » eux diminuent tandis que les « cinquantenaires » augmentent légèrement.

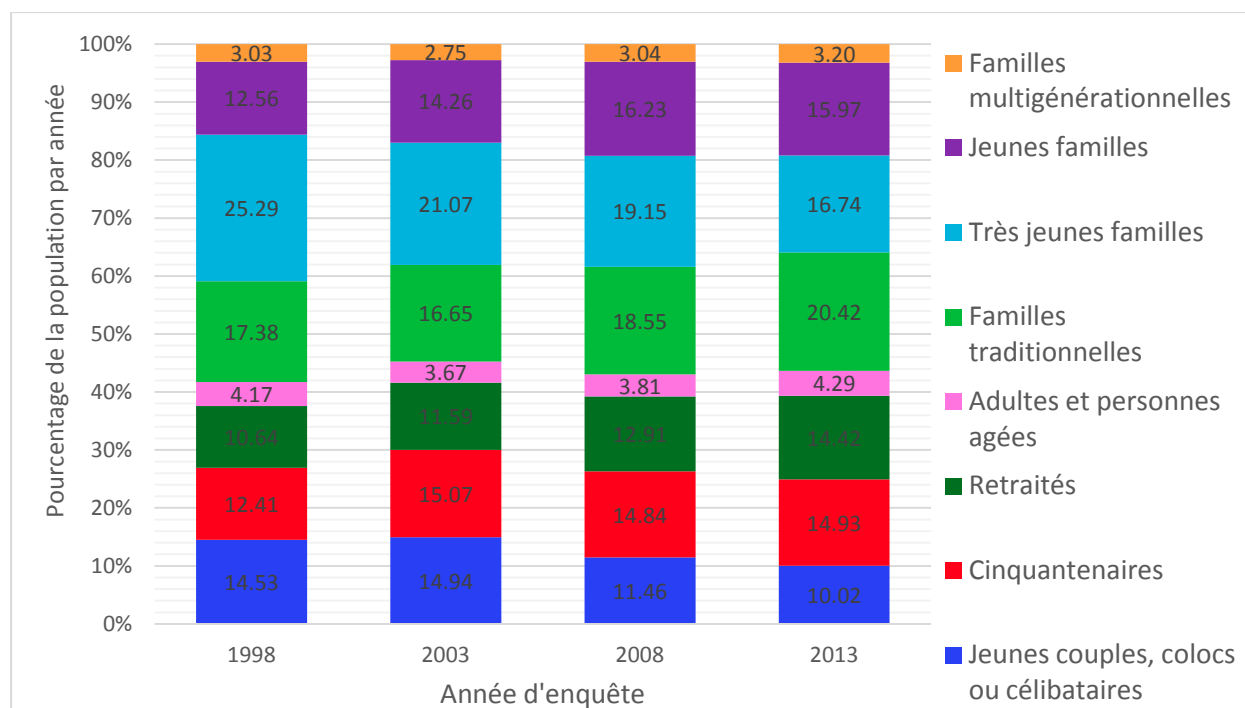


Figure 5.11 : Appartenance aux types de ménages à travers le temps

Il peut être pertinent de regarder si ces évolutions d'appartenance à un type de ménage concernent toutes les catégories de personnes ou seulement quelques-unes. La Figure 5.12 montre l'évolution de l'appartenance à un type de ménage au travers des années, pour les 40-49 ans (environ 17% de la population interrogée), tandis que la Figure 5.13 montre cette évolution pour les 70-79 ans (environ 7% de la population interrogée). Ainsi, les variations sont très faibles pour les 70-79 ans : la majorité de cette catégorie de personne appartient au type « retraités » et l'augmentation du nombre de personnes dans ce dernier groupe est totalement expliquée par l'augmentation du nombre de personnes âgées de 70 à 79 ans. Au contraire, chez les 40-49 ans, les variations sont plus grandes et plus nombreuses. L'échange entre les « jeunes familles » et les « très jeunes familles » est de nouveau observable. Pour ce qui est des « familles traditionnelles », les 40-49 ans appartiennent davantage à ce type de ménage en 2008 que dans les années précédentes, mais la part diminue de 2008 à 2013. Ainsi l'augmentation de la part des « familles traditionnelles » sur la population totale entre 2008 et 2013 doit être causée par une autre catégorie de personnes.

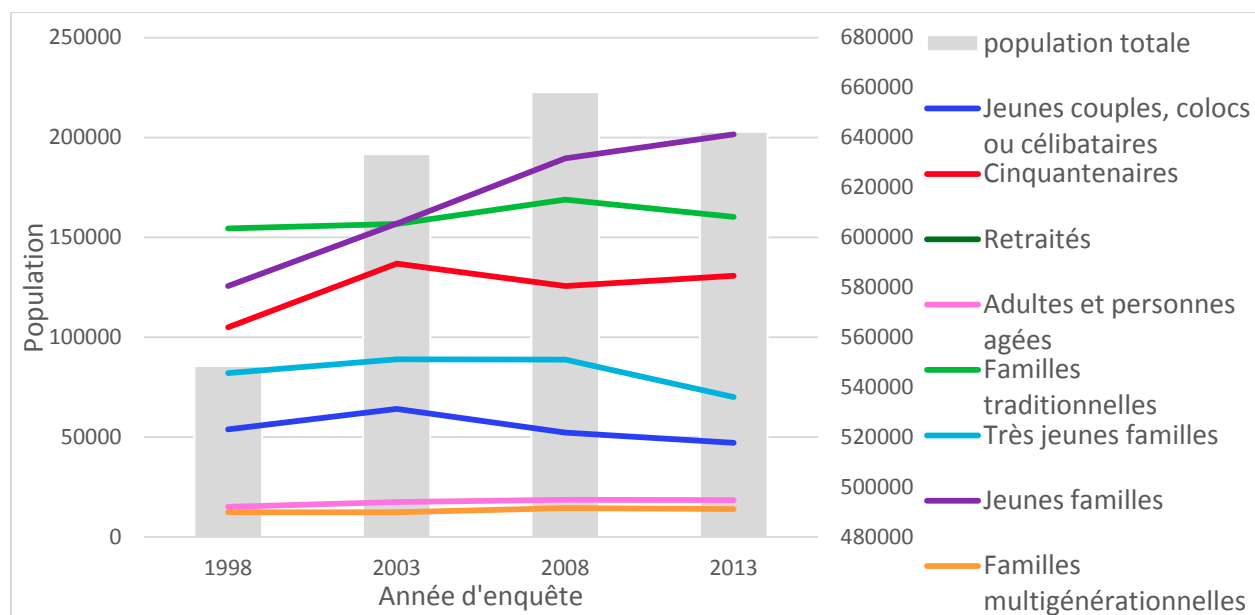


Figure 5.12 : Appartenance aux types de ménages à travers le temps pour les 40-49 ans

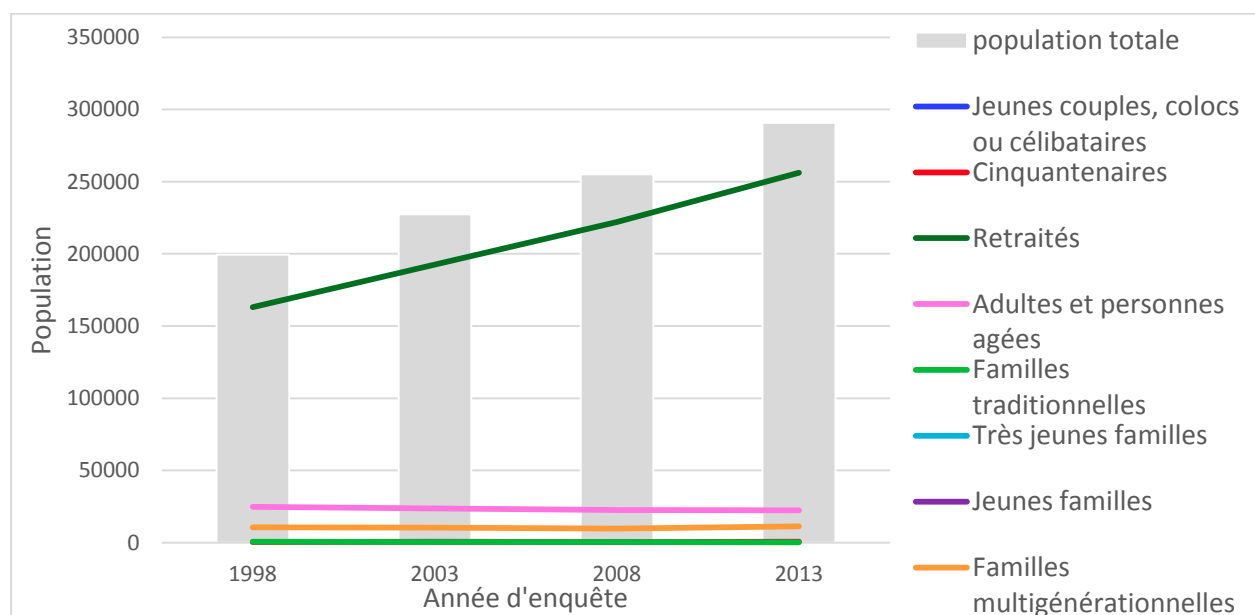


Figure 5.13 : Appartenance aux types de ménages à travers le temps pour les 70-79 ans

5.2.2 Différences de comportements de déplacement d'un type de ménage à un autre

Outre la transférabilité temporelle, il est important que la typologie retenue soit pertinente, c'est-à-dire que les comportements de déplacement doivent être différents d'un groupe à l'autre. Des

différences sont observables pour la plupart des indicateurs de déplacements de la personne. La section suivante présente les valeurs moyennes des cinq indicateurs de déplacement pour chaque type de ménage de la typologie effectuée sur l'ensemble des enquêtes. Celle-ci est située dans la méthodologie globale dans la Figure 5.14

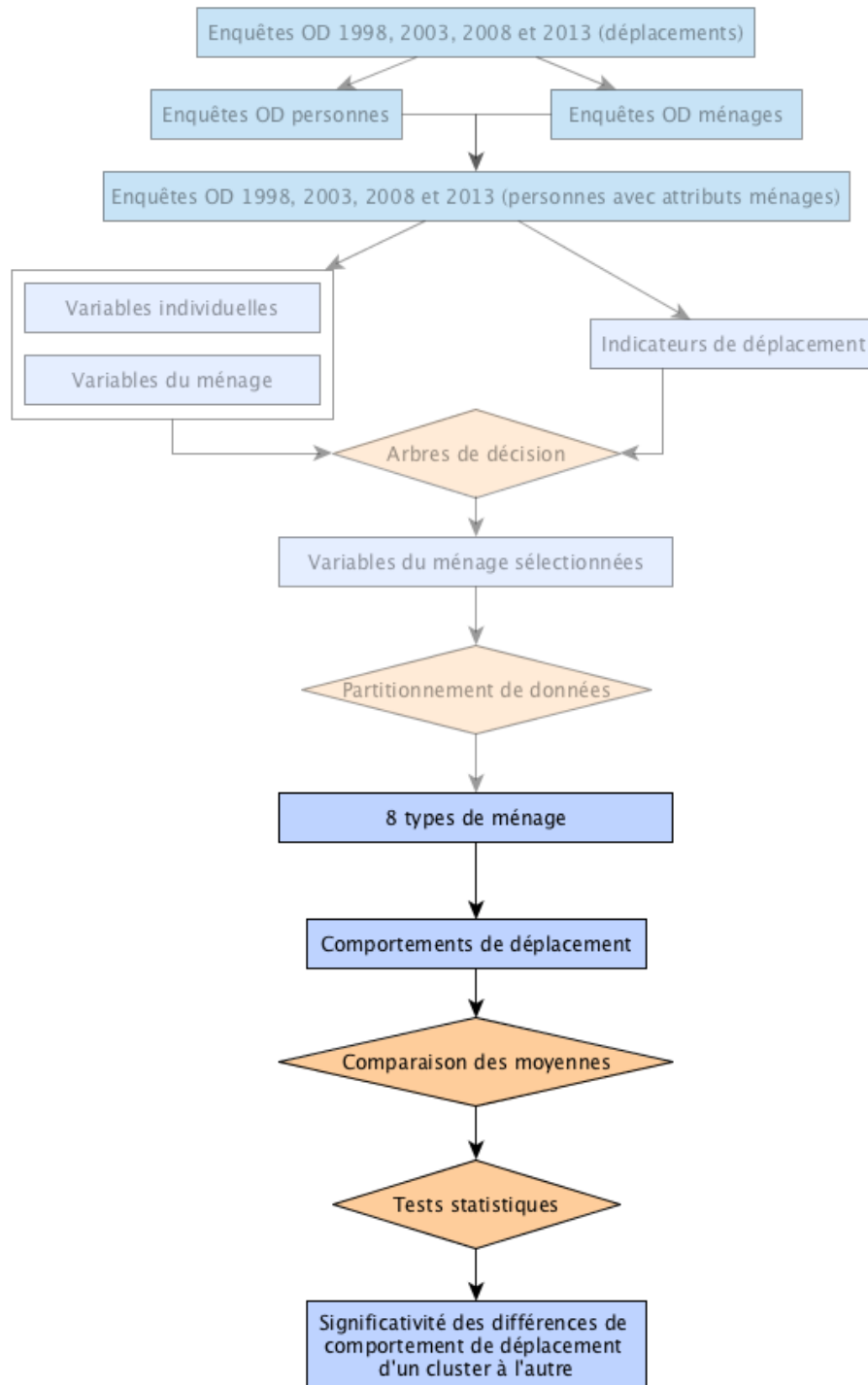


Figure 5.14 : Position de la section suivante dans le schéma méthodologique

5.2.2.1 Mobilité

En ce qui concerne le fait d'être mobile ou non, seuls les types de ménages 3 et 4 présentent des parts élevées (environ 40 et 33% respectivement) de personnes non mobiles. En effet ce sont les types « retraités » et « adultes et personnes âgées » et ce n'est pas étonnant, car il est bien connu que les personnes âgées se déplacent moins que les actifs. Les groupes 1, 5, 6 et 7 ont tous des parts de personnes mobiles supérieures à 85%. Ces derniers ont tous des moyennes d'âge inférieures à 40 ans et sont la plupart du temps des ménages au sein desquels des enfants sont présents. Leur mobilité est donc plus importante.

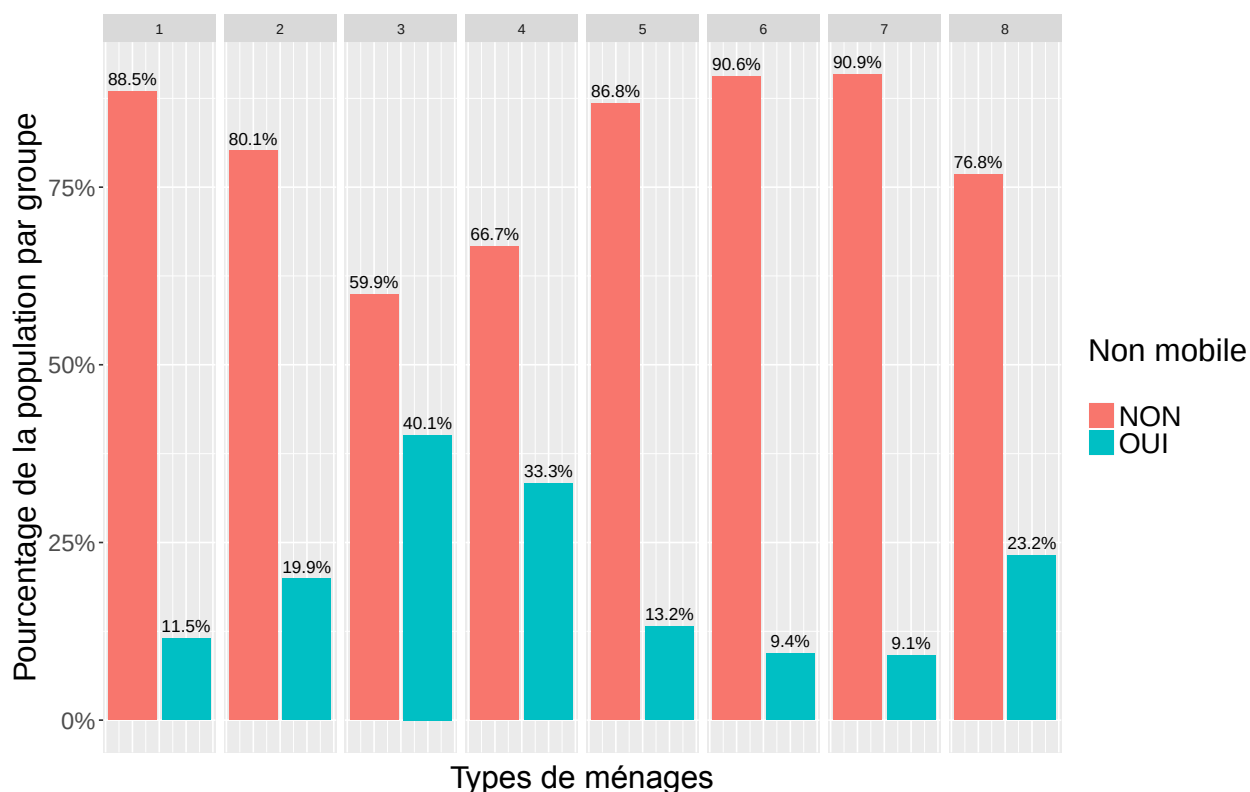


Figure 5.15 : Part de personnes mobiles et non mobiles dans les 8 types de ménage

5.2.2.2 Nombre de déplacements

Les différences de nombre de déplacements d'un type de ménage à un autre sont semblables à celles observées pour la mobilité. Le groupe 3 peut être différencié du groupe 4, car la part la plus importante des personnes appartenant à ce groupe n'effectue aucun déplacement tandis que pour le groupe 4, la majorité des personnes effectue un déplacement par jour (sans compter les

retours). Seuls les groupes 6 et 7, à savoir les jeunes et très jeunes familles, ont une part de personnes effectuant cinq déplacements et plus non négligeable.

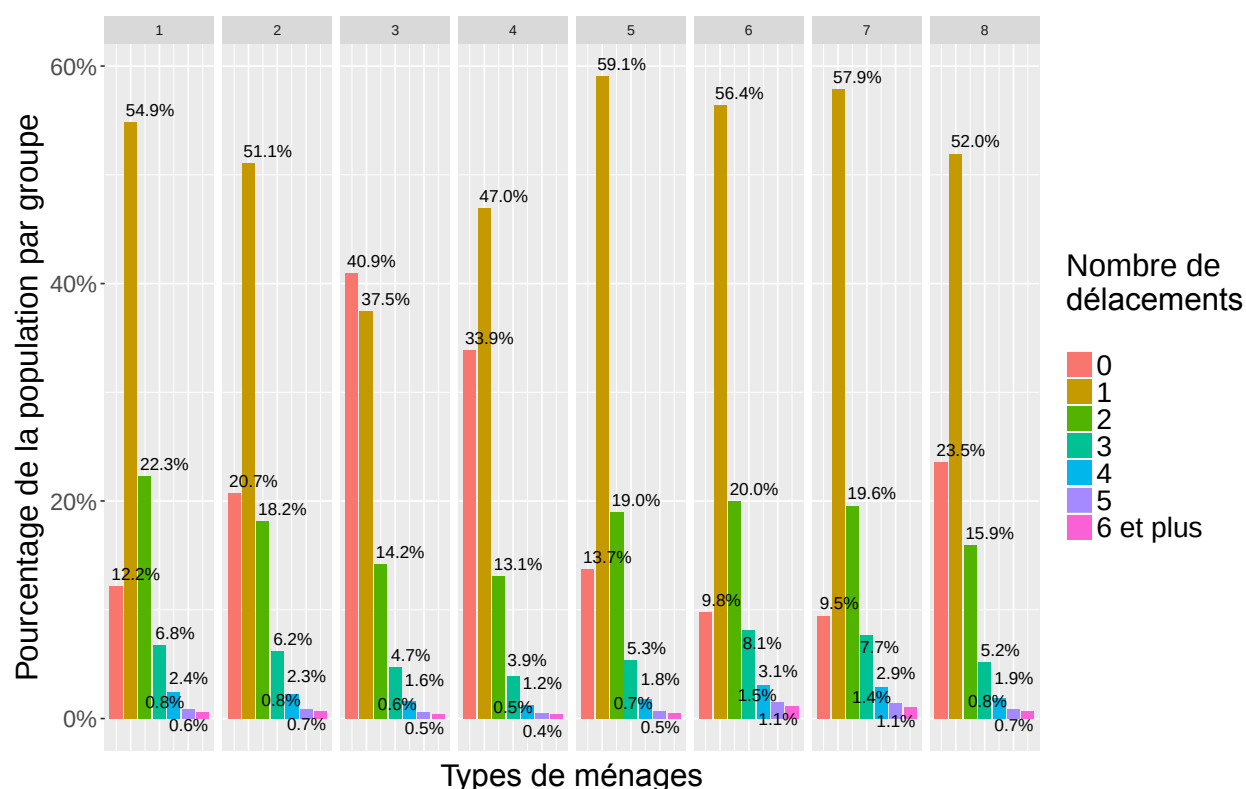


Figure 5.16 : Nombre de déplacements (excluant les retours) par personne dans les 8 types de ménage

5.2.2.3 Part de déplacements en pointe matinale

En ce qui concerne la part de déplacements en pointe matinale, une fois de plus, les groupes 3 et 4 se démarquent (très faible part de déplacement en pointe AM pour la majorité des personnes de ces groupes). Les groupes 6 et 7 sont ceux qui se déplacent le plus en pointe AM. Ceci s'explique par le fait que ce sont des ménages composés le plus souvent de travailleurs avec enfants dont les horaires sont généralement moins flexibles. Les « jeunes couples, colocs ou célibataires » sont nombreux à ne pas se déplacer en pointe matinale, il en est de même pour les « cinquantenaires ». Ainsi, il est possible que ce soit davantage la présence d'enfants qui force les déplacements en pointe matinale que le fait d'être travailleur.

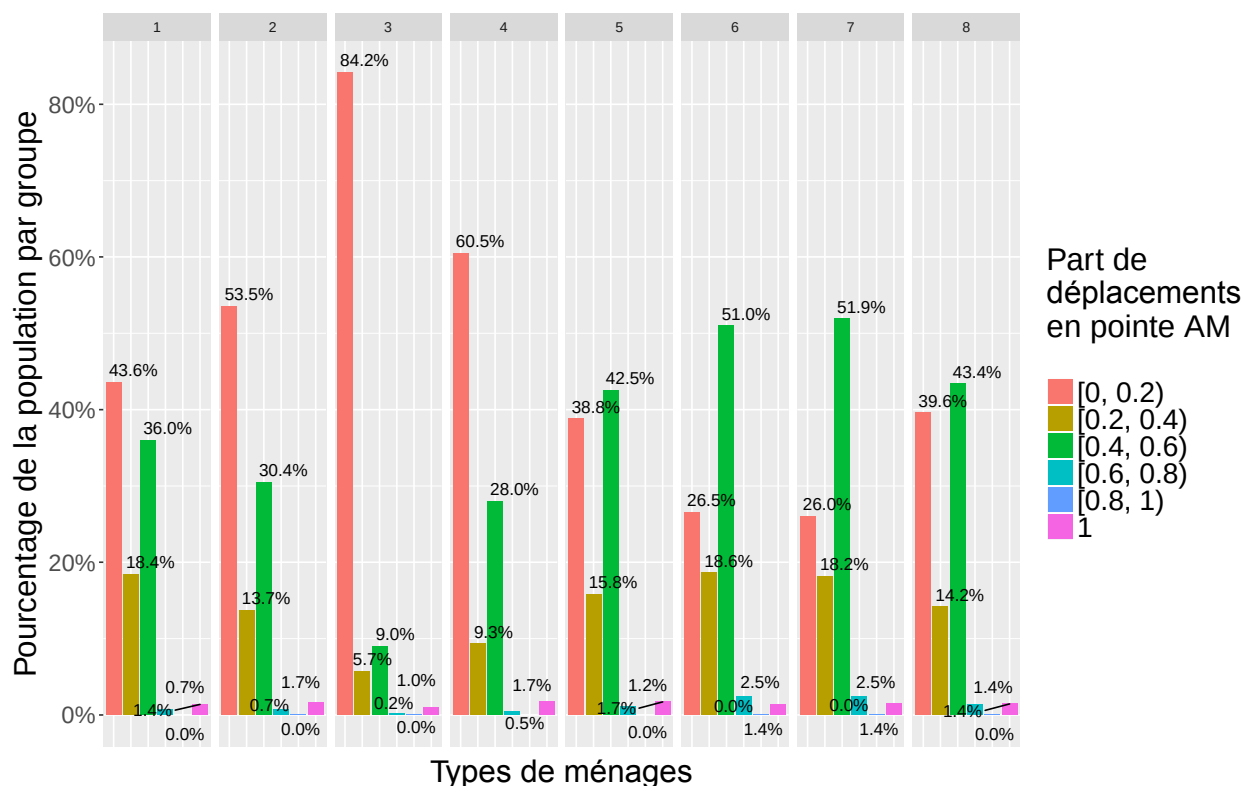


Figure 5.17 : Part de déplacements en pointe matinale dans les 8 types de ménage

5.2.2.4 Part de déplacements en auto

Pour la part de déplacements en automobile (conducteur ou passager), un point important à remarquer est que les personnes effectuant tous leurs déplacements en auto sont majoritaires dans tous les groupes. Seuls les groupes 1, 6, 7 et 8 présentent des parts importantes de personnes utilisant d'autres modes. L'âge est donc un des déterminants de l'utilisation de la voiture. Pour les groupes 6, 7 et 8, qui correspondent aux familles, la part élevée de personnes n'utilisant pas tout le temps l'automobile vient également du fait que, dans les ménages nombreux, les personnes doivent se partager la ou les automobiles.

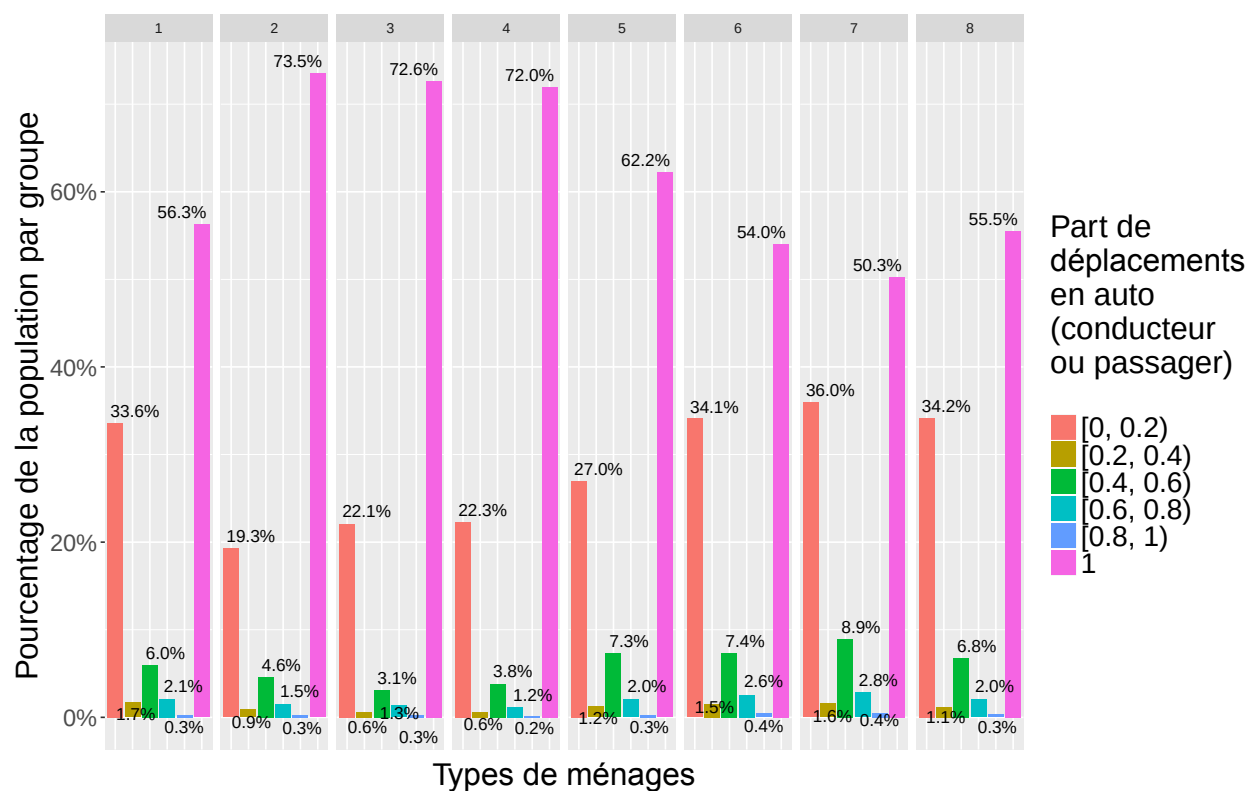


Figure 5.18 : Part de déplacements en auto dans les 8 types de ménage

5.2.2.5 Chaîne complexe

Finalement, pour le fait de faire au moins une chaîne complexe dans sa journée ou pas, les différences ne sont pas très visibles d'un type de ménage à un autre. Le groupe « retraités » présente la part de personnes effectuant au moins une chaîne complexe la plus importante.

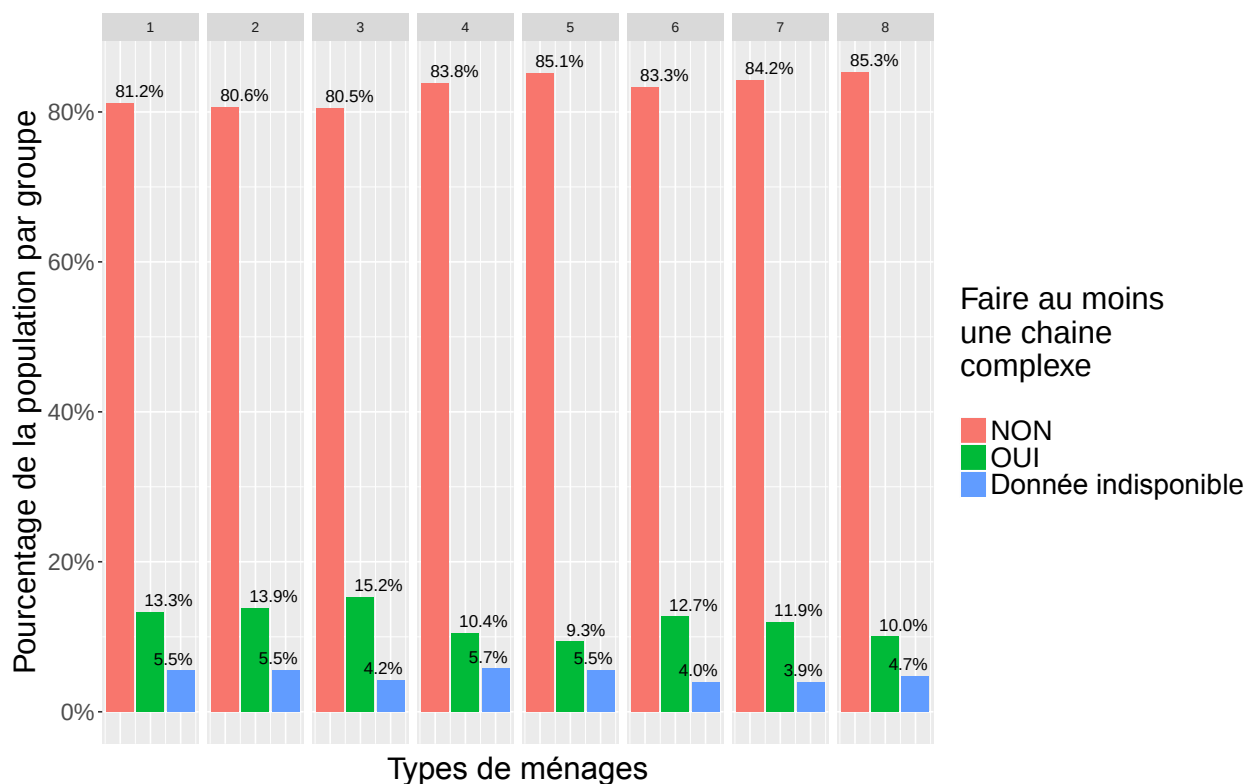


Figure 5.19 : Part de personnes faisant une chaîne complexe ou plus dans les 8 types de ménage

Ces représentations graphiques permettent de déceler des différences de comportement de mobilité entre les groupes et même d'affecter certains comportements à un type de ménage. Cependant, il est important, avant de statuer que les types de ménages sont corrélés avec les déplacements de la personne, de vérifier que ces différences sont bien significatives.

5.2.3 Significativité des différences de comportement de déplacement : tests statistiques

Pour confirmer les différences de déplacements d'un type de ménage à un autre observées ci-dessus, des tests statistiques sont appliqués afin de déterminer si les indicateurs de déplacement sont effectivement différents d'un type de ménage à un autre. Pour cela, des comparaisons de la moyenne de chaque indicateur de déplacement seront d'abord exécutées, puis les distributions seront confrontées. Ici, huit échantillons seront comparés pour cinq indicateurs dont les valeurs peuvent être continues (nombre de déplacements, part de déplacements en automobile et part de déplacements en pointe matinale) ou bien binaires (être mobile ou non et faire une chaîne complexe ou pas).

5.2.3.1 Comparaison des moyennes

Pour la comparaison de la moyenne, le test de Tukey a été appliqué sur les cinq indicateurs, comme décrit dans la méthodologie. Les Tableau 5.6, Tableau 5.7, Tableau 5.8, Tableau 5.9 et Tableau 5.10 rassemblent les valeurs-p des tests de Tukey pour chaque indicateur. Pour rappel, une faible valeur-p (inférieure à 0,01, 0,05 ou 0,1 suivant le degré de significativité), implique que l'hypothèse nulle peut être rejetée à savoir qu'il est impossible de considérer que les groupes ont la même moyenne (ou bien distribution dans la section suivante). Il est alors dit que les différences sont statistiquement significatives. Dans ces tableaux, lorsque p est inférieure à $2e-16$, la valeur 0 est appliquée.

Tableau 5.6 : Valeurs-p pour le test de Tukey sur la mobilité

	Jeunes couples, colocs ou célibataires	Cinquantenaires	Retraités	Adultes et personnes âgées	Familles traditionnelles	Très jeunes familles	Jeunes familles
Cinquantenaires	0						
Retraités	0	0					
Adultes et personnes âgées	0	0	0				
Familles traditionnelles	5,85E-14	0	0	0			
Très jeunes familles	0	0	0	0	0		
Jeunes familles	0	0	0	0	0	0,669628	
Familles multigénérationnelles	0	0	0	0	0	0	0

Tableau 5.7 : Valeurs-p pour le test de Tukey sur le nombre de déplacements

	Jeunes couples, colocs ou célibataires	Cinquantennaires	Retraités	Adultes et personnes âgées	Familles traditionnelles	Très jeunes familles	Jeunes familles
Cinquantennaires	0						
Retraités	0	0					
Adultes et personnes âgées	0	0	0,028885				
Familles traditionnelles	0	3,97E-09	0	0			
Très jeunes familles	0	0	0	0	0		
Jeunes familles	0	0	0	0	0	3,19E-05	
Familles multigénérationnelles	0	6,94E-14	0	0	0	0	0

Tableau 5.8 : Valeurs-p pour le test de Tukey sur la part de déplacements en pointe matinale

	Jeunes couples, colocs ou célibataires	Cinquantennaires	Retraités	Adultes et personnes âgées	Familles traditionnelles	Très jeunes familles	Jeunes familles
Cinquantennaires	0						
Retraités	0	0					
Adultes et personnes âgées	0	0	0				
Familles traditionnelles	0	0	0	0			
Très jeunes familles	0	0	0	0	5,68E-06		
Jeunes familles	0	0	0	0	5,78E-06	0,99996	
Familles multigénérationnelles	0	0	0	0	0	0	0

Tableau 5.9 : Valeurs-p pour le test de Tukey sur la part de déplacements en automobile

	Jeunes couples, colocs ou célibataires	Cinquantennaires	Retraités	Adultes et personnes âgées	Familles traditionnelles	Très jeunes familles	Jeunes familles
Cinquantennaires	0						
Retraités	0	0					
Adultes et personnes âgées	0	0	0				
Familles traditionnelles	0	0	0	0			
Très jeunes familles	0	0	0	0	0		
Jeunes familles	0	0	0	0	0	7,78E-14	
Familles multigénérationnelles	0	0	0	0	0	0	0

Tableau 5.10 : Valeurs-p pour le test de Tukey sur le booléen faire une chaîne complexe ou non

	Jeunes couples, colocs ou célibataires	Cinquantennaires	Retraités	Adultes et personnes âgées	Familles traditionnelles	Très jeunes familles	Jeunes familles
Cinquantennaires	0						
Retraités	0	0					
Adultes et personnes âgées	0	0	0				
Familles traditionnelles	0,975772	0	0	0			
Très jeunes familles	0	0	0	0	0		
Jeunes familles	0	0	0	0	0	0,003881	
Familles multigénérationnelles	0	0,147926	0	0	0	0	0

Tableau 5.12 : Valeurs p pour le test de Kolmogorov-Smirnov sur la mobilité (suite)

Groupes	H50_55	F50_55	H55_60	F55_60	H60_65	F60_65	H65_70	F65_70	H70_75	F70_75	H75_80	F75_80	H80_85	F80_85	H85_90	F85_90	H90_95	F90_95	H95_100	F95_100
1 & 2																				
1 & 3																				
1 & 4																				
1 & 5																				
1 & 6																				
1 & 7																				
1 & 8																				
2 & 3																				
2 & 4																				
2 & 5																				
2 & 6																				
2 & 7																				
2 & 8																				
3 & 4																				
3 & 5																				
3 & 6																				
3 & 7																				
3 & 8																				
4 & 5																				
4 & 6																				
4 & 7																				
4 & 8																				
5 & 6																				
5 & 7																				
5 & 8																				
6 & 7																				
6 & 8																				
7 & 8																				

Ces valeurs sont calculées pour chaque couple de types de ménages pour toutes les catégories genre*âge (catégories d'âge de 5 ans). Des nuances de couleur sont utilisées pour représenter la valeur de p. Plus le vert est foncé, plus la valeur de p est faible et donc plus les différences entre les groupes sont significatives. Ainsi, en plus de noter que les différences entre certains couples sont plus prononcées que pour d'autres, l'âge et le genre ont également une importance. Pour l'ensemble des indicateurs étudiés, les différences de comportements de déplacement sont en effet plus significatives pour les catégories d'âge entre 15 et 65 ans. Aussi, les différences sont généralement plus significatives chez les femmes. Pour les 50-55 ans, chez les hommes comme chez les femmes et pour tous les indicateurs, les différences sont très peu significatives.

En conclusion, même si les tests de significativité ne permettent pas de dire que les indicateurs de déplacement sont différents pour l'intégralité des groupes, cela est bien le cas pour la majorité

d'entre eux. Cette analyse permet également de noter que pour certains groupes d'âge et de genre, les différences sont significatives entre deux groupes, tandis qu'elles ne le sont pas pour d'autres catégories. Ainsi, l'affectation des personnes à un type de ménages devra être plus rigoureuse pour les catégories de personnes pour lesquelles les différences sont les plus significatives.

5.3 Autres typologies testées

Comme cela a été mentionné dans la section 3.5, d'autres méthodes de regroupement des données ont été testées. Les différences concernent la normalisation des données avant d'appliquer l'algorithme des k-moyennes. Cependant, ces méthodes n'ont pas été retenues dans les présents travaux. Plusieurs raisons sont à l'origine de ce choix. Premièrement, que ce soit la normalisation « min-max » ou bien le « z-score », aucun des deux ne procurait une aussi bonne séparation des données que sans normalisation. C'est-à-dire que plusieurs groupes possédaient des moyennes similaires pour certaines variables. Aussi, le fait de donner moins d'importance aux variables du nombre de personnes de moins de 12 ans ou de 76 à 84 ans dans le ménage a été validé. En effet, la présence de ces catégories de personnes influe sur l'âge moyen ou bien l'amplitude maximale des âges du ménage. Ainsi, le fait de leur donner moins d'importance dans l'élaboration des groupes compense cette redondance. Enfin, le modèle d'affectation qui est décrit dans le chapitre suivant était beaucoup moins performant sur les groupements de données faits avec normalisation. L'utilisation de la normalisation « min-max » ou bien du « z-score » n'aurait donc pas permis une bonne prévision de la répartition des types de ménages dans la population, ou du moins, elle aurait été moins bonne qu'avec le regroupement de données effectué sans normalisation. La Figure 5.20 illustre un exemple de typologie avec la normalisation « min-max ».

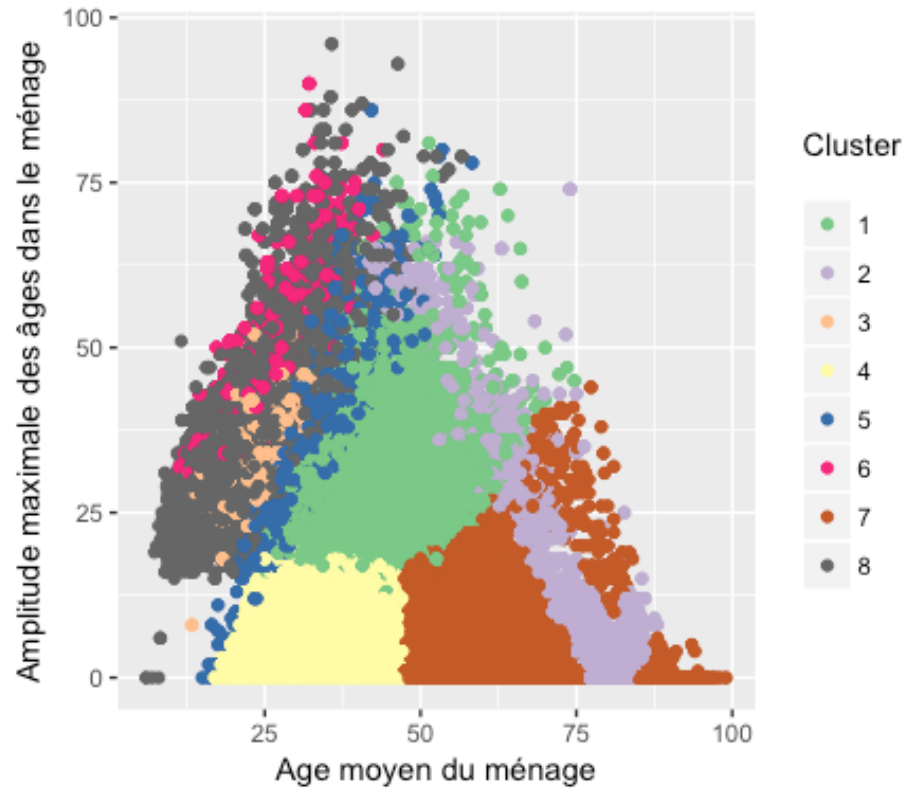


Figure 5.20 : Exemple de regroupement de données avec la normalisation "min-max"

CHAPITRE 6 MODÈLE D'AFFECTATION DES PERSONNES À UN GROUPE DE MÉNAGES

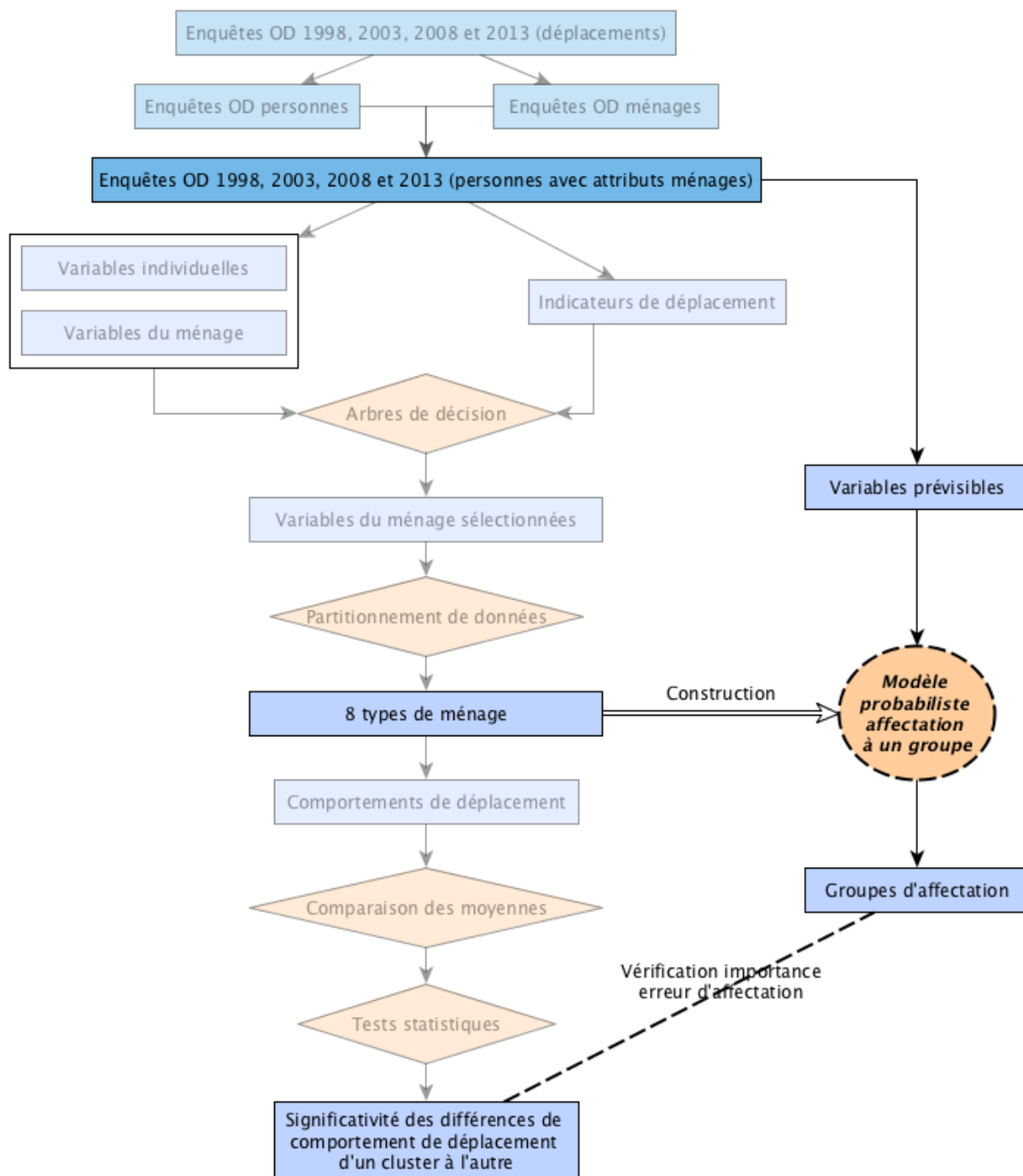


Figure 6.1 : Position de la section suivante dans le schéma méthodologique

Ultime étape de ce projet, la création d'un modèle probabiliste d'affectation va permettre d'associer chaque personne à un des huit types de ménages présentés ci-dessus, suivant ses caractéristiques personnelles. Cette étape est située dans la méthodologie globale dans la Figure 6.1.

6.1 Formation du modèle

Pour créer le modèle, la volonté ici est de ne s'appuyer que sur les variables qui sont facilement prévisibles dans les projections de la population, à savoir l'âge, le genre et la région du domicile. Comme précisé dans la méthodologie, le modèle logit multinomial est choisi ici. Plusieurs étapes se sont succédé pour construire le modèle. Tout d'abord, un modèle a été fait pour chaque année d'enquête séparément. Finalement, comme pour la création de la typologie, un modèle commun à toutes les années d'enquête est préféré. Celui-ci intègre donc une variable pour l'année d'enquête (« nombre d'années depuis la source », la source étant 1998). Le modèle logit multinomial créé pour l'affectation des personnes à un type de ménage prend donc comme variables entrantes l'âge de la personne (variable continue, nombre entier de 0 à 99), son genre (deux variables catégoriques : « homme » et « femme »), la région de son domicile (huit variables catégoriques pour les huit régions présentes dans l'enquête OD) et enfin l'année d'enquête de laquelle fait partie la personne (variable continue, nombre entier de 0 à 15).

6.2 Description du modèle

Les risques relatifs, c'est-à-dire les rapports de la probabilité d'appartenir à un type de ménage sur la probabilité d'appartenir au type de ménage de référence, à savoir les « jeunes couples, colocs ou célibataires » sont donnés dans le Tableau 6.1. Les étoiles à côté du risque relatif représentent la significativité des variables. Par exemple, dans ce modèle, un accroissement d'une unité de l'âge donne lieu à une probabilité d'appartenir aux « Retraités » 1,492 fois plus grande que la probabilité d'appartenir au groupe de référence. À l'inverse, toujours pour un accroissement d'une unité de l'âge, la probabilité d'appartenir aux très jeunes familles est égale à 0,970 fois la probabilité d'appartenir au groupe de référence. En plus des effets de l'âge sur l'appartenance à un type de ménage, ce modèle permet de faire ressortir d'autres aspects tels que l'appartenance à un type de ménage suivant la région du domicile. Ainsi, en ce qui concerne les groupes de familles (groupes 5, 6 et 7), la probabilité d'appartenir à ces types de ménages est

largement plus grande (par rapport au groupe de référence) lorsque la région du domicile est autre que le centre ou le centre-ville.

Tableau 6.1 : Risques relatifs du modèle logit multinomial

	Variable dépendante :						
	Cinquantennaires	Retraités	Adultes et personnes âgées	Familles traditionnelles	Très jeunes familles	Jeunes familles	Familles multigénérationnelles
Homme	0.109***	0.001***	0.041***	0.773***	1.454***	1.052***	0.298***
Femme	0.137***	0.001***	0.053***	0.856***	1.759***	1.185***	0.392***
Âge	1.170***	1.492***	1.209***	1.028***	0.970***	0.988***	1.054***
Région 1	0.340***	0.101***	0.270***	0.406***	0.410***	0.508***	0.426***
Région 2	0.336***	0.099***	0.293***	0.415***	0.485***	0.550***	0.430***
Région 3	0.615***	0.175***	0.602***	1.048***	1.151***	1.141***	0.932***
Région 4	0.556***	0.170***	0.573***	1.274***	1.176***	1.517***	1.220***
Région 5	0.719***	0.204***	0.499***	1.201***	1.370***	1.174***	0.754***
Région 6	0.680***	0.201***	0.546***	1.309***	1.583***	1.320***	0.971
Région 7	0.923***	0.269***	0.552***	1.406***	2.179***	1.317***	0.948**
Région 8	0.854***	0.260***	0.531***	1.330***	2.009***	1.264***	0.806***
Nombre d'années depuis 1998	1.044***	1.058***	1.040***	1.054***	1.011***	1.069***	1.065***
Constante	0.015***	0.00000*	0.002***	0.662***	2.558***	1.247***	0.117***
Akaike Inf. Crit.	1,84E+06	1,84E+06	1,84E+06	1,84E+06	1,84E+06	1,84E+06	1,84E+06
Note :							*p<0.1; **p<0.05; ***p<0.01

6.2.1 Résultats et performance du modèle

Pour évaluer le modèle, la matrice de confusion de ce dernier peut être créée. Cependant, la fonction *predict* de R précédemment utilisée pour les matrices de confusion se contente de prendre le type de ménage qui présente la plus grande probabilité pour un enregistrement donné. Elle ne peut donc pas être utilisée ici. Ainsi, dix simulations ont été faites en prenant un nombre aléatoire entre 0 et 1 pour affecter un type de ménage à chaque observation de la base de données. Les matrices de confusion de ces dix simulations sont toutes très similaires, l'une d'elles est représentée en Figure 6.2 et Tableau 6.2. Le détail des matrices de confusion pour les

dix simulations est présenté en [annexe J](#). Il est possible de voir grâce à ces matrices que le modèle présente des lacunes. Effectivement, certains groupes sont très peu prédits, c'est le cas des « Adultes et personnes âgées » et les « Familles multigénérationnelles ». Pourtant, même si elles sont peu, des observations de ces groupes sont faites en réalité. Au contraire, on observe sur la matrice de confusion que les groupes correspondant à des familles (« Familles traditionnelles » et « Très jeunes familles » particulièrement) sont largement trop prédits. Aussi, parmi ces mêmes types de ménages, de nombreuses personnes ont été interverties, c'est-à-dire qu'elles sont prédites dans le groupe 6 alors qu'elles appartiennent en réalité au groupe 5 ou bien au groupe 7 ou inversement. Il est possible que cela soit dû au fait que le modèle fait difficilement la différence entre ces deux groupes. Probablement qu'une variable entrante supplémentaire donnant des informations sur la composition du ménage permettrait d'améliorer cela.

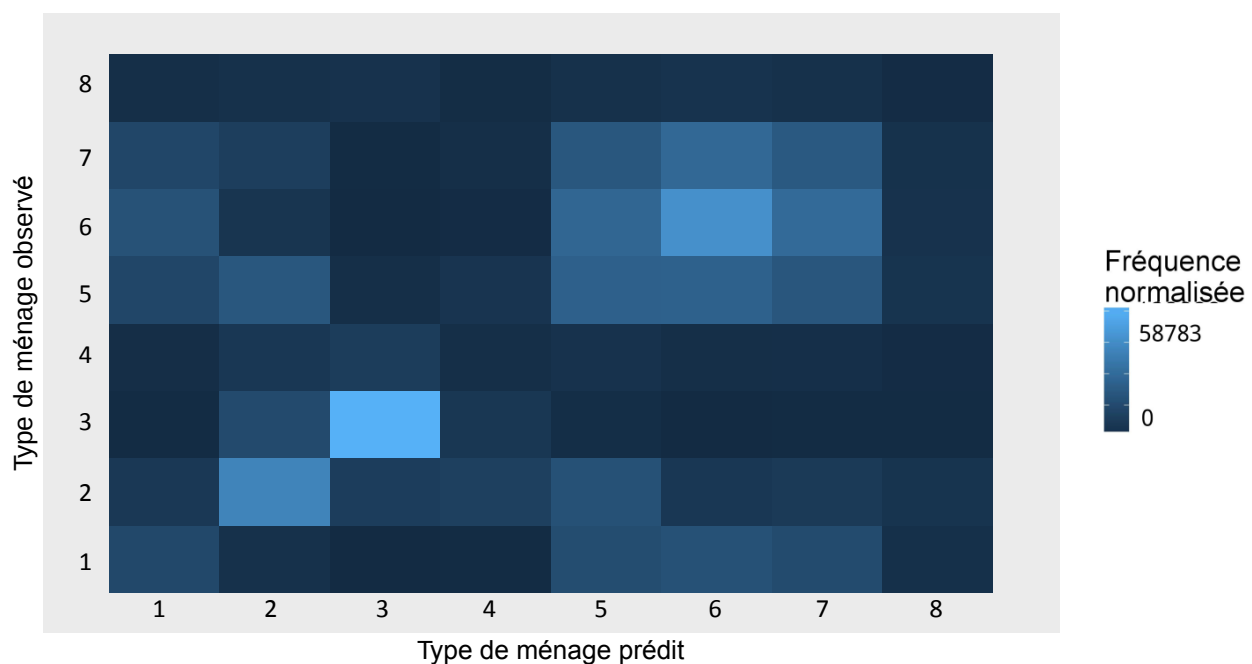


Figure 6.2 : Matrice de confusion du modèle logit multinomial (simulation 1)

Tableau 6.2 : Matrice de confusion du modèle logit multinomial (simulation 1)

	1	2	3	4	5	6	7	8	Total prédits
1	13 946	6 487	331	1 660	13 211	18 564	13 311	1 997	69 507
2	2 838	40 373	15 034	6 157	20 736	4 776	9 492	2 978	102 384
3	5	9 081	58 783	8 709	1 829	5	277	3 633	82 322
4	416	10 331	5 796	1 814	4 727	679	1 878	865	26 506
5	16 230	18 220	1 413	3 543	25 071	27 485	20 908	3 040	115 910
6	18 094	6 081	179	1 960	25 591	45 618	28 560	3 997	130 080
7	15 283	7 383	329	2 104	20 501	29 264	21 676	3 038	99 578
8	2 438	4 317	485	729	4 332	3 659	3 414	590	19 964
Total observés	69 250	102 273	82 350	26 676	115 998	130 050	99 516	20 138	646 251

6.2.2 Utilisation du modèle

Le but du modèle d'affectation créé est de savoir à quel type de ménage les personnes vont appartenir pour pouvoir prédire leur comportement de déplacement en fonction des caractéristiques du ménage dans lequel elles vivent. Ainsi, même si, comme le montre la section 6.2.1, le modèle d'affectation ne permet pas d'affecter rigoureusement chaque personne au type de ménage auquel elle appartient réellement, ce dernier peut donner les probabilités de la personne d'appartenir à chaque groupe. Ces probabilités sont donc utilisées pour prévoir une répartition de la population au sein des différents types de ménages et à travers les années. Celle-ci est donnée en Figure 6.3. Cette répartition de la population ne permet certes pas de savoir quelle personne est rattachée à quel type de ménage, mais on obtient grâce à elle une estimation du nombre de personnes qui appartiennent respectivement à chaque type de ménages. Finalement, il est donc possible d'estimer le comportement de mobilité de ces personnes en prenant en compte les caractéristiques de leur ménage.

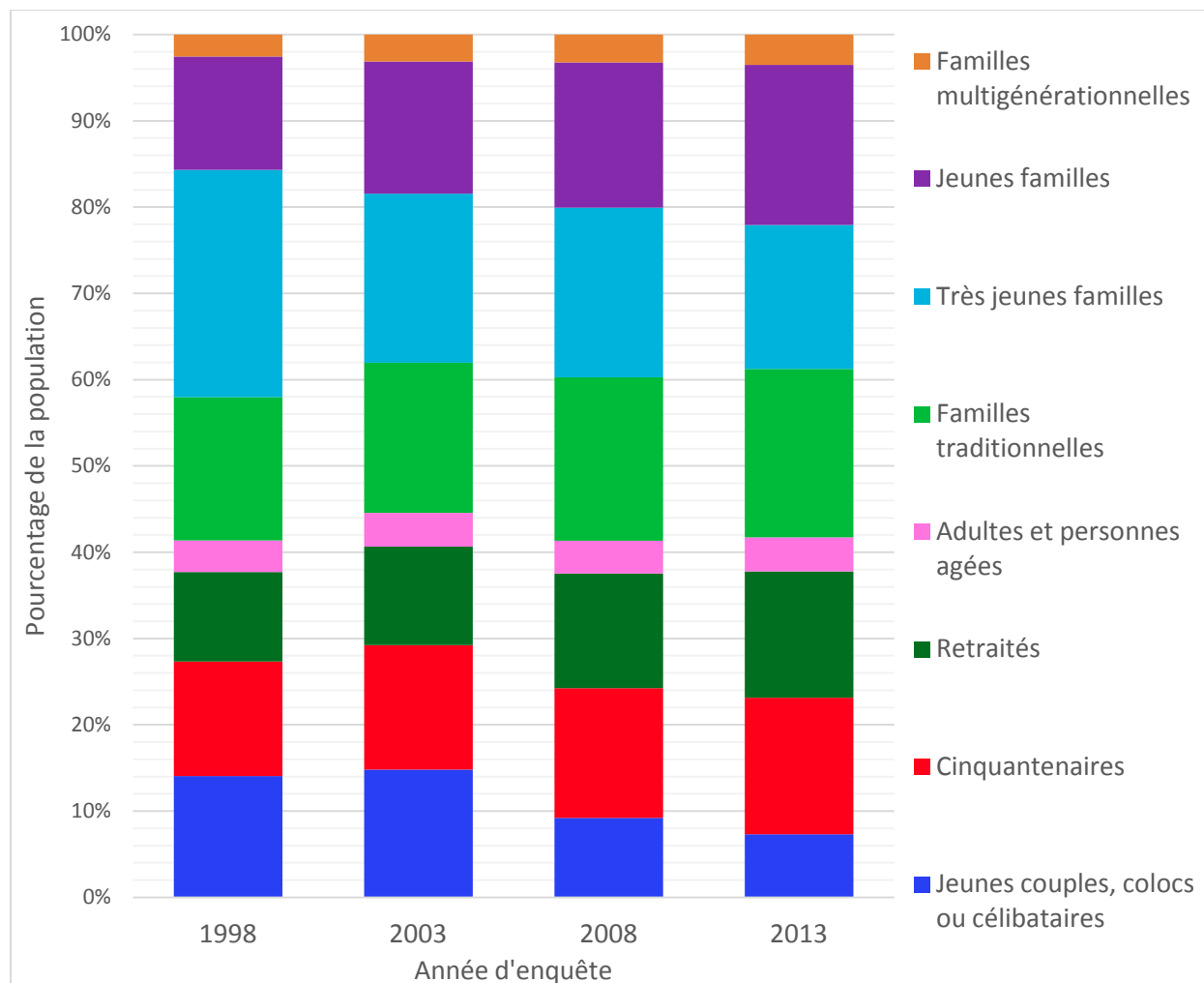


Figure 6.3 : Répartition de la population dans les types de ménages prévue à la,ide du modèle

6.2.3 Évolution des répartitions

Les probabilités d'appartenir à un type de ménage pour les années 1998, 2003, 2008 et 2013 ont été utilisées afin d'estimer l'évolution des parts de chaque type de ménage à travers le temps. Ainsi, à partir du pourcentage de la population appartenant à un type de ménage donné pour les 4 années d'enquête, une régression est appliquée pour estimer le pourcentage de la population dans ce même groupe pour le futur (2021 dans ce cas). La Figure 6.4 illustre les régressions pour chaque type de ménage.

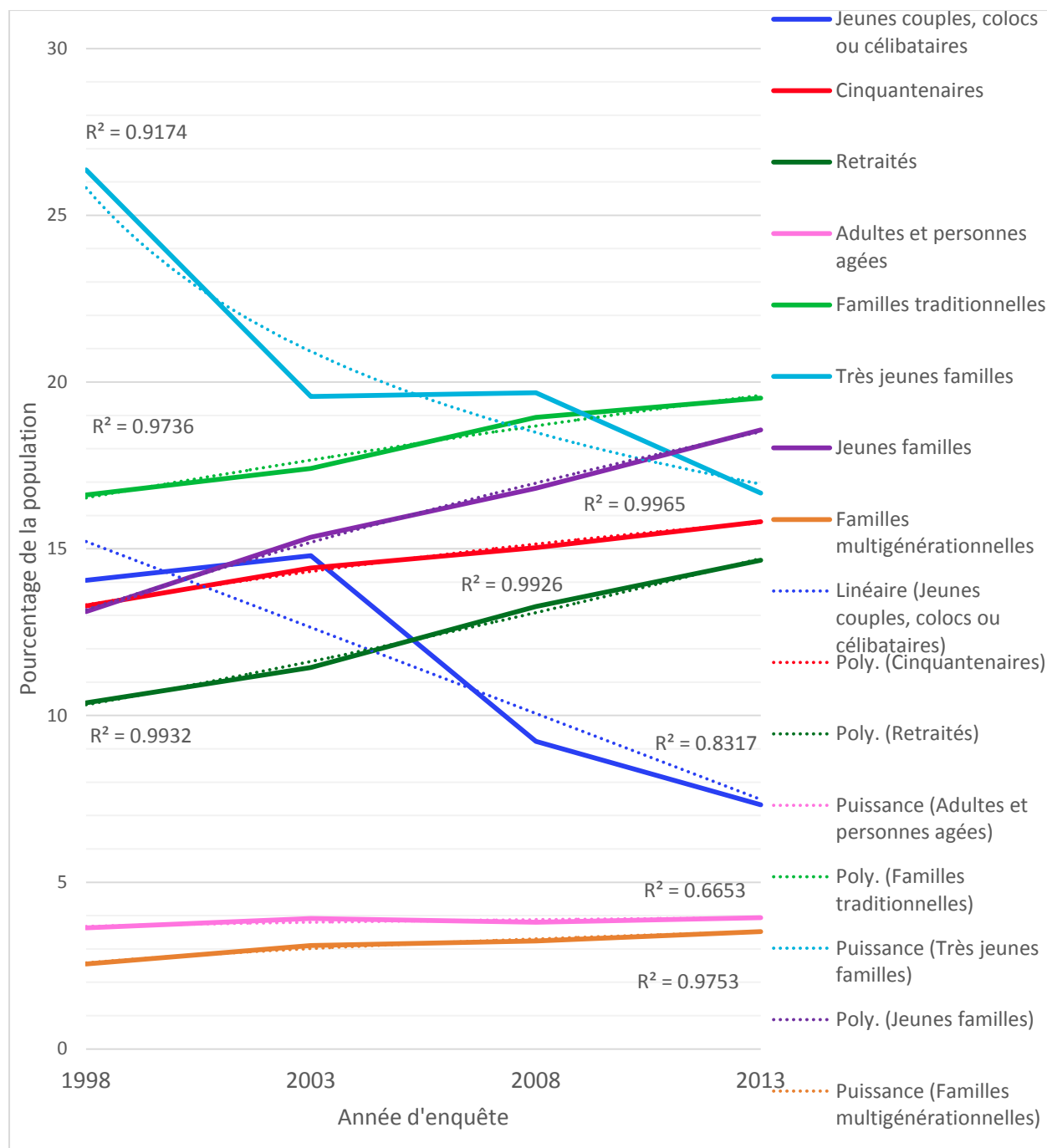


Figure 6.4 : Courbes de tendance de l'évolution des parts de chaque type de ménage

Ensuite, avec ces régressions, la répartition de la population dans les types de ménages est prévue pour 2021. Celle-ci est représentée en Figure 6.5. Les équations représentant l'évolution de la part de chaque type de ménage sont ajustées afin d'obtenir un total de 100%.

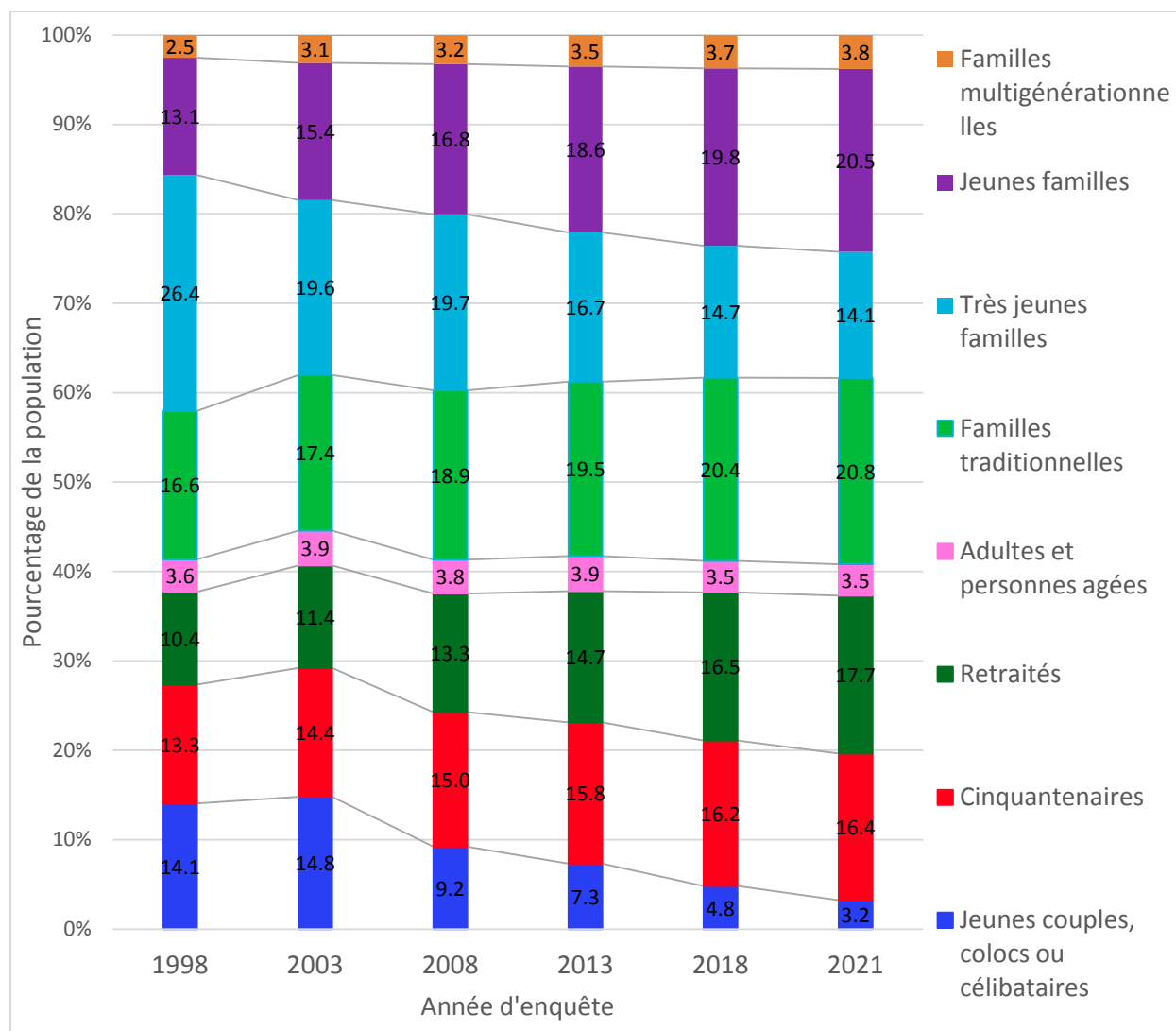


Figure 6.5 : Prédiction de l'évolution de la répartition des types de ménages pour 2021

6.2.4 Prévisions pour 2021 et 2031

Le modèle a été appliqué pour prévoir la répartition de la population en 2021 et 2031. Les facteurs prévisionnels démographiques issus des *Projections démographiques scénario MO2017_01* ont été utilisés. Aucune évolution des proportions de chaque type de ménage vu au-dessus n'a été prise en compte ici, les évolutions de la répartition de la population à l'intérieur des groupes se basent donc essentiellement sur les évolutions démographiques. À partir de la base de données de l'enquête OD de 2013, le modèle logit multinomial est appliqué à la population avec les facteurs d'expansion prévus pour 2021 et pour 2031. Ainsi, la répartition des personnes dans les ménages, obtenue pour ces deux années, est donnée en Figure 6.6. Il est

possible d'observer l'augmentation progressive de la part des « retraités », contrairement aux parts des groupes de familles (« familles traditionnelles », « très jeunes familles » et « jeunes familles ») qui diminuent. Ceci illustre parfaitement le vieillissement de la population prévu pour les années à venir. Si la part des groupes « retraités » et « cinquantenaires » obtenue dans ce cas est plutôt similaire à celle obtenue en section 6.2.3, ce n'est pas le cas pour les groupes « jeunes couples, colocs ou célibataires » ainsi que pour les groupes de famille. Ceci confirme donc que le modèle a du mal à différencier les types de familles, comme mentionné en section 6.2.1. Aussi, pour ces types de ménage, il serait pertinent de mêler évolutions démographiques et les régressions linéaires effectuées dans la section 6.2.3.

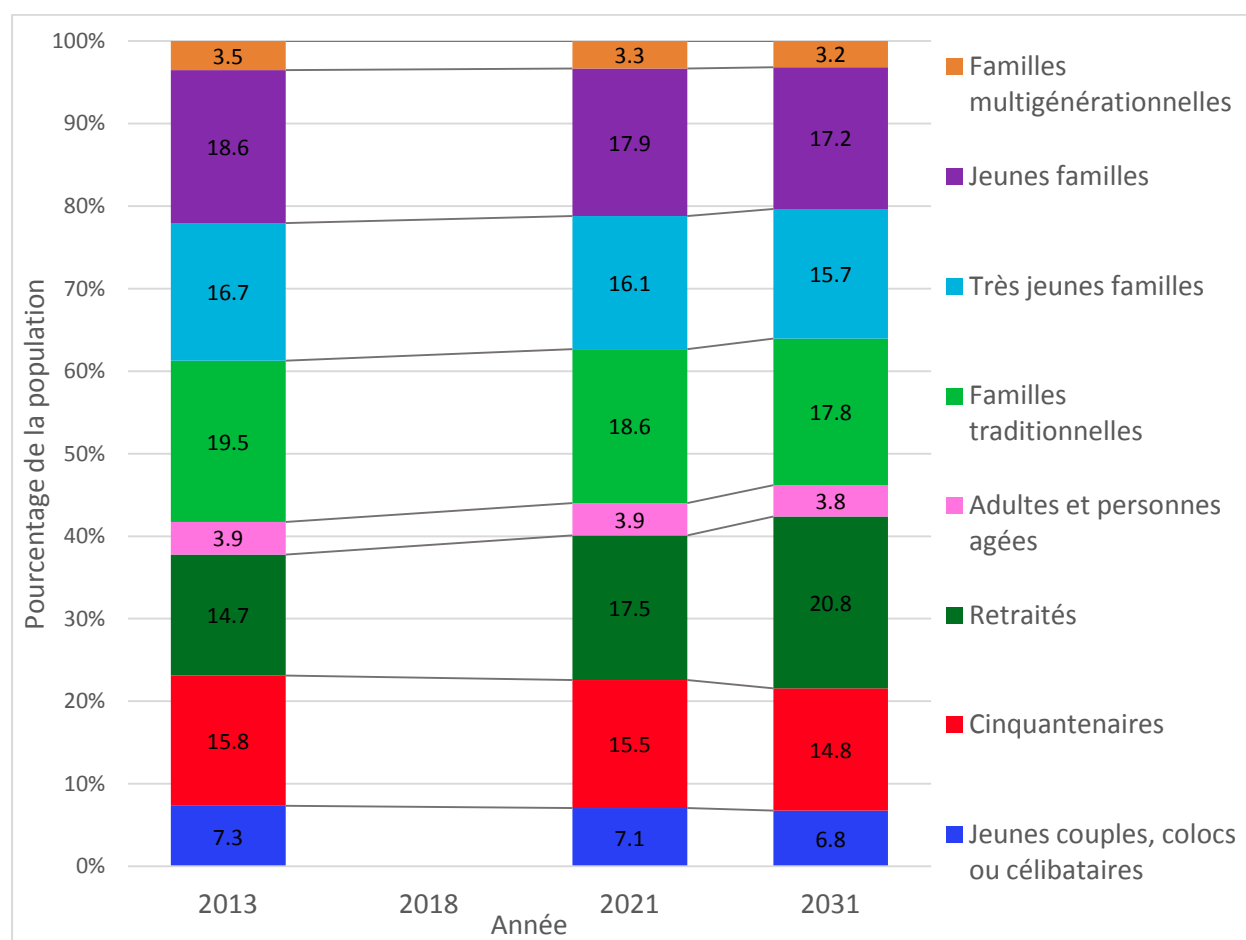


Figure 6.6 : Préviction de la répartition de la population dans les différents types de ménages en 2021 et 2031

Plusieurs scénarios d'évolution de la population pourraient également être effectués, de cette manière, les scénarios « extrêmes » de demande en transport pourraient être envisagés. Par

exemple, on pourrait imaginer un scénario où les naissances sont très nombreuses en banlieue, mais pas dans le centre de Montréal et un autre où l'inverse serait prévu.

CHAPITRE 7 CONCLUSION

Le présent chapitre, dernier de ce mémoire, permet de conclure ce projet de recherche dont le but principal était de créer une typologie de ménages, afin d'intégrer les caractéristiques de ces derniers à la prévision des comportements individuels de déplacement. Les recherches se sont donc concentrées sur la problématique suivante : la prise en compte de l'entourage de la personne dans l'analyse de ses déplacements et dans la prévision de ses déplacements futurs. La première section de ce chapitre rappelle les principaux résultats des recherches, puis une deuxième section traite des limitations rencontrées au cours de ce projet de recherche. Enfin, une dernière section permet d'imaginer quelles trajectoires pourraient être empruntées pour aller plus loin sur ce sujet. Toutes les recherches de ce projet ont été effectuées sur les enquêtes Origine-Destination de la Grande région de Montréal des années 1998, 2003, 2008 et 2013.

7.1 Synthèse de la recherche et contributions

Le premier chapitre de ce mémoire constitue une revue de la littérature sur les principaux sujets reliés à ce projet de recherche, à savoir : les évolutions démographiques, les évolutions de tendances de mobilité, les liens qui peuvent être formés entre les deux et enfin les typologies de ménages déjà observées dans la littérature. Certaines tendances relativement connues ont d'abord été expliquées. Celles-ci sont relatives à la population, comme le vieillissement de la population, la diminution de la taille des ménages et les changements dans la répartition spatiale de ces derniers, ou bien relatives à la mobilité, comme la diminution de la non mobilité ou bien l'augmentation de la part modale des transports en commun. Puis, parmi les caractéristiques de la personne et du ménage ayant une influence sur les comportements de déplacement, il a été remarqué que l'âge des personnes qui composent le ménage a un réel impact sur les déplacements d'une personne, notamment la présence de membres « dépendants » (ce sont principalement les enfants en bas âge et les personnes âgées ne pouvant plus effectuer tous leurs déplacements seuls). Enfin, dans les typologies de ménages relevées dans la littérature, certaines sont pertinentes, mais la plupart se basent sur la notion de couple et représentent uniquement les ménages « traditionnels ». Or, de nos jours, de nombreux autres types de ménages apparaissent et deviennent de plus en plus nombreux (ménages monoparentaux, colocations de personnes de plus en plus âgées, etc.). Cette étape de revue de littérature a donc permis de confirmer le besoin

de créer une nouvelle forme de typologie de ménages ainsi que d'orienter le choix des variables à prendre en compte dans cette dernière.

Ensuite, une sélection plus fine des variables à considérer dans l'élaboration de la typologie a été faite dans le Chapitre 4. Celle-ci est effectuée grâce à des arbres de décision et de régression. Pour évaluer les différentes variables de la personne et du ménage, cinq indicateurs de déplacement ont été choisis pour représenter la mobilité d'une personne, ce sont : le fait d'être mobile ou non, le nombre de déplacements effectués sur une journée (sans compter les retours), la part de déplacements effectués en pointe matinale, la part de déplacements effectués en automobile et le fait de faire une chaîne complexe dans sa journée ou pas. Ainsi, cinq arbres ont été formés par année d'enquête. L'analyse de ces derniers permet de tirer les variables ayant le plus d'impact sur les indicateurs de déplacements. Le choix final des variables (comportant quelques subjectivités présentées dans la section 4.2.2) s'est porté sur : l'âge moyen du ménage, l'amplitude maximale des âges dans le ménage, le nombre de personnes dans le ménage, la présence d'enfants de moins de 12 ans et la présence de personnes âgées de 76 à 84 ans.

Une fois la sélection des caractéristiques du ménage ayant le plus d'influence sur les comportements de mobilité individuels effectuée, la typologie de ménage peut être créée dans le Chapitre 5. Celle-ci se base sur un algorithme de fouille de données, les k-moyennes. La typologie retenue est faite à partir d'une base de données regroupant toutes les années d'enquête, pour une meilleure transférabilité temporelle, et présente huit groupes, basés sur les caractéristiques du ménage. Les types de ménages obtenus sont donc principalement séparés selon l'âge des personnes qui les composent et peuvent être énumérés sous les noms suivants : les « jeunes couples, colocs ou célibataires », les « cinquantenaires », les « retraités », les « adultes et personnes âgées », les « familles traditionnelles », les « très jeunes familles », les « jeunes familles » et les « familles multigénérationnelles ». L'évolution de l'appartenance des personnes à un type de ménage est étudiée pour vérifier la transférabilité temporelle de la typologie et les comportements de déplacement des personnes appartenant à chaque groupe sont étudiés pour en vérifier la pertinence.

Enfin, un modèle d'affectation des personnes à un type de ménage est créé dans le Chapitre 6. La force de celui-ci, qui peut également être une faiblesse, est qu'il ne prend en variables entrantes que certaines variables prédites et entrantes dans le modèle de prévision, à savoir l'âge, le genre

et la région du domicile de la personne. Ceci en fait une force, car peu d'informations sont nécessaires pour prédire l'affectation d'une personne à un groupe, mais une faiblesse également, car, à cause du peu de variables entrantes, le modèle est moins performant. Ainsi, le modèle obtenu ne permet pas de faire une prédiction rigoureuse du type de ménage auquel la personne est affectée par rapport au type de ménage auquel elle appartient réellement. Cependant, ce dernier permet d'obtenir une répartition d'une population synthétique au sein des différents groupes de ménages.

7.2 Limitations

Il est important de revenir sur les quelques limites rencontrées tout au long de ce projet, qu'elles soient relatives à l'utilisation des données ou bien à celle des outils techniques.

Premièrement, ce travail de maîtrise s'insère dans un projet global, à savoir la prise en compte des caractéristiques du ménage et des systèmes d'activités pour la synthétisation de la population et de sa mobilité. Ainsi, le projet doit tenir compte des éléments qui l'entourent. En l'occurrence, les enquêtes OD et les variables utilisées par le MTMDET dans le processus actuel de prévision de la demande d'un côté et l'objet nécessaire à la synthétisation de la population et de sa mobilité de l'autre. Ce dernier prend la forme d'une typologie de ménages qui permettrait de passer d'une prévision de personnes totalement énumérées à une prévision de ménages totalement énumérés.

En ce qui concerne les données utilisées, la typologie de ménages et le modèle d'affectation créés dans ce projet ont été construits à partir d'enquêtes OD présentant des facteurs de pondération. Or ces facteurs de pondération présentent une limite à évaluer rigoureusement la composition de la population et surtout sa répartition géographique précise. D'où l'un des buts du projet global de synthétiser la population à partir de l'échantillon observé pour obtenir une population totalement énumérée avec une meilleure variabilité des comportements et des attributs.

Ensuite, la composition des ménages varie beaucoup au cours des années comme cela a été mentionné dans la revue de littérature. De nouveaux types de ménages apparaissent, d'autres disparaissent, la récurrence des différents types de ménages varie... Tous ces changements sont dus à des évolutions politiques, culturelles et économiques qu'il est bien difficile de prévoir.

Aussi, des limites sont observées dans les méthodes employées. Par exemple, le choix des cinq indicateurs pour représenter la mobilité des personnes est loin d'être exhaustif et pourrait être jugé subjectif. De même, la détermination des variables à intégrer à la typologie se fait sur des critères choisis par l'auteure. Aussi, sur les variables utilisées, le fait de ne considérer que le premier mode emprunté constitue une limite par exemple.

Enfin, des limites dans l'application des algorithmes d'arbre de décision, des k-moyennes ou bien du modèle logit ont été observées, principalement en raison de la taille des échantillons utilisés. Pour les arbres de décision par exemple, l'exemple du nombre de déplacements montre que l'arbre de décision prédit surtout les nombres de déplacements « couramment observés » (un ou deux déplacements sans retour dans ce cas). Les alternatives peu présentes dans les observations sont souvent « oubliées » par les algorithmes ou les modèles (c'est également le cas pour les groupes de ménages 4 et 8 de la typologie dans le modèle d'affectation).

7.3 Recherches futures et perspectives

Avec des données aussi riches que celles des enquêtes OD, le champ des possibles est très large. Ainsi, les analyses détaillées dans ce mémoire pourraient être étoffées en utilisant davantage de données de l'enquête OD, notamment pour les indicateurs de déplacement de la personne. Ainsi, le mode utilisé pourrait être pris en compte, de même que les distances parcourues, le temps hors domicile, etc.

Quelques améliorations pourraient encore être apportées à ce modèle d'affectation. Tout d'abord, le statut de la personne pourrait être ajouté aux variables entrantes du modèle logit multinomial, sous forme de variables catégoriques. En effet, le statut de la personne peut être prédit par le MTMDET. Des hypothèses sont posées sur « l'évolution future de la proportion d'un statut pour une combinaison d'un âge, sexe et région du domicile » (MTMDET, 2017b). La population est catégorisée selon quatre statuts (travailleur, étudiant, autre mobile et non mobile). Une modification pourrait également être apportée dans le modèle sur la variable « Nombre d'années depuis la source ». Celle-ci pourrait être convertie en plusieurs variables catégoriques. En effet, dans l'état actuel du modèle, cette dernière prend beaucoup d'importance au fur et à mesure que l'on s'éloigne de 1998. De plus, l'importance de la variable d'année d'enquête varie suivant l'année source choisie. Aussi, la notion de gradient spatial pourrait davantage être prise en

compte concernant le modèle d'affectation. Par exemple, un modèle différent pourrait être développé pour chaque région du domicile afin de faire prendre plus d'importance à cette variable et de mieux voir les effets sur la probabilité d'appartenir à un type de ménage des autres variables en fonction du lieu du domicile. Il faudrait également que l'utilisation du modèle soit optimisée. La prise en compte de l'évolution des coefficients sur les quatre précédentes années devrait être mêlée à l'application du modèle à une population prédite pour 2021 et 2031. Ainsi, la prévision de la répartition de la population dans les ménages tiendrait à la fois compte des évolutions démographiques mais aussi des tendances des parts de chaque type de ménages observées dans le passé. Également cela pourrait être fait pour plusieurs scénarios d'évolution de la population. Enfin, il faudrait illustrer la contribution de la création d'une typologie de ménages. Pour cela, un indicateur de déplacement devrait être choisi et le comportement moyen de la population devrait être comparé au comportement moyen de chaque type de ménage. Ainsi, la comparaison des tendances avec et sans prise en compte de la structure du ménage pourrait venir confirmer l'impact de cette dernière sur la prévision des déplacements.

Finalement, comme cela a été mentionné précédemment, cette recherche s'insère dans un projet plus global dont le but est de synthétiser la population et sa mobilité en prenant en compte les caractéristiques du ménage et les systèmes d'activités. Ainsi les principales perspectives de recherche pour donner suite à ce mémoire seront au nombre de deux. Premièrement, il s'agira de construire une méthode de désagrégation de la population à utiliser sur l'enquête OD afin d'obtenir des personnes complètement énumérées (dont le facteur de pondération est égal à un). Une fois cette étape effectuée, la typologie de ménages, construite dans le présent mémoire, pourra être utilisée, de même que le modèle d'affectation. Ceci permettra d'obtenir une énumération de ménages (dont le facteur de pondération sera d'un également). Deuxièmement, un modèle basé activités pourra être appliqué aux ménages désagrégés afin d'obtenir les déplacements, eux aussi complètement énumérés. Ainsi, appliquée à une prévision de la population, cette procédure permettra d'obtenir la prévision des déplacements futurs, désagrégés, en prenant en compte à la fois la structure du ménage et les systèmes d'activités.

CHAPITRE 8 BIBLIOGRAPHIE

- Agarwal, A. (2004). A comparison of weekend and weekday travel behavior characteristics in urban areas.
- Agence Métropolitaine de Transport. (2010). Enquête Origine-Destination 2013, La mobilité des personnes dans la région de Montréal–Faits Saillants: Montréal. Consulté le.
- Alsnih, R., & Hensher, D. A. (2003). The mobility and accessibility expectations of seniors in an aging population. *Transportation Research Part A: Policy and Practice*, 37(10), 903-916.
- Arbonville, D., & Bonvalet, C. (1992). Conditions de logement et types de ménages. Résultats de l'enquête-logement 1988. *Population (french edition)*, 1294-1306.
- Bernard, A., Seguin, A.-M., Bussiere, Y., & Polacchini, A. (2000). *Household structure and mobility patterns of women in OD surveys: Methods and results based on the case studies of Montreal and Paris*. Communication présentée à Women's Travel Issues Second National Conference.
- Bernardo, C., Paleti, R., Hoklas, M., & Bhat, C. (2015). An empirical investigation into the time-use and activity patterns of dual-earner couples with and without young children. *Transportation Research Part A: Policy and Practice*, 76, 71-91.
- Bhat, C. R., Goulias, K. G., Pendyala, R. M., Paleti, R., Sidharthan, R., Schmitt, L., & Hu, H.-H. (2013). A household-level activity pattern generation model with an application for Southern California. *Transportation*, 40(5), 1063-1086.
- Bhat, C. R., & Singh, S. K. (2000). A comprehensive daily activity-travel generation model system for workers. *Transportation Research Part A: Policy and Practice*, 34(1), 1-22.
- Bonnel, P. (2002). *Prévision de la demande de transport*. (Université Lumière-Lyon II).
- Bonnel, P., Le Nir, M., & Nicolas, J.-P. (1994). Les enquêtes déplacements urbains. Réflexions méthodologiques sur les enquêtes ménages et les enquêtes régionales origine destination canadiennes.
- Bradley, M., & Vovsha, P. (2005). A model for joint choice of daily activity pattern types of household members. *Transportation*, 32(5), 545-571.
- Breiman, J., & Friedman, R. (1984). Olsen and CJ Stone, "Classification And Regression Trees", Monterey. CA: Wadsworth.
- Brownson, R. C., & Boehmer, T. K. (2004). Patterns and trends in physical activity, occupation, transportation, land use, and sedentary behaviors. *TRB Special Report*, 282.
- Buchanan, N., & Barnett, R. (2006). Peripheral residential relocation and travel pattern change. *Urban policy and research*, 24(2), 217-236.
- Bussière, Y. (1990). Effet du vieillissement démographique sur la demande de transport dans la région métropolitaine de Montréal, 1986-2011. *Cahiers québécois de démographie*, 19(2), 325-351.

- Cerema Direction Technique Territoires et Ville. (2001). Définition d'une typologie de la mobilité adaptée à la modélisation de la demande de transport dans les agglomérations françaises. Tiré de <http://www.innovations-transports.fr/Definition-d-une-typologie-de-la>
- Chan, A., Saito, Y., & Robine, J.-M. (2016). International perspectives on summary measures of population health in an aging world: SAGE Publications Sage CA: Los Angeles, CA.
- Collin, J.-P., & Mongeau, J. (1992). Quelques aspects démographiques de l'étalement urbain à Montréal de 1971 à 1991 et leurs implications pour la gestion de l'agglomération. *Cahiers québécois de démographie*, 21(2), 5-30.
- Communauté métropolitaine de Montréal. (2015). PROJECTIONS DÉMOGRAPHIQUES : 221000 nouveaux ménages attendus dans le Grand Montréal pour la période 2016-2031. *Perspective Grand Montréal*(29). Tiré de http://cmm.qc.ca/fileadmin/user_upload/periodique/29_Perspective.pdf
- Coulombel, N. (2010). Residential choice and household behavior: State of the Art. *Ecole Normale Supérieure de Cachan*.
- Courson, J.-P., & de Saboulin, M. (1985). Ménages et familles: vers de nouveaux modes de vie? *Economie et statistique*, 175(1), 3-20.
- Crane, R. (2007). Is there a quiet revolution in women's travel? Revisiting the gender gap in commuting. *Journal of the American planning association*, 73(3), 298-316.
- Desgraupes, B. (2012). *Notes de cours Anova*. Université Paris Ouest Nanterre La Défense. Tiré de <http://bdesgraupes.pagesperso-orange.fr/UPX/Master1/anovaExample.pdf>
- Desgraupes, B. (2014). *Notes de cours Méthodes statistiques*. Université Paris Ouest Nanterre La Défense. Tiré de http://bdesgraupes.pagesperso-orange.fr/UPX/L2/MethStats_seance_11_doc.pdf
- Désilets, S. (2012). *Tendances liées à la composition, la motorisation et la localisation spatiale des ménages*. (École Polytechnique de Montréal).
- Fan, Y. (2017). Household structure and gender differences in travel time: spouse/partner presence, parenthood, and breadwinner status. *Transportation*, 44(2), 271-291.
- Gaube, V., & Remesch, A. (2013). Impact of urban planning on household's residential decisions: An agent-based simulation model for Vienna. *Environmental Modelling & Software*, 45, 92-103.
- Hanson, S. (1982). The determinants of daily travel-activity patterns: relative location and sociodemographic factors. *Urban Geography*, 3(3), 179-202.
- Hensher, D. A., & Reyes, A. J. (2000). Trip chaining as a barrier to the propensity to use public transport. *Transportation*, 27(4), 341-361.
- Hildebrand, E. D. (2003). Dimensions in elderly travel behaviour: A simplified activity-based model using lifestyle clusters. *Transportation*, 30(3), 285-306.
- Hu, X., Wang, J., & Wang, L. (2013). Understanding the travel behavior of elderly people in the developing country: a case study of Changchun, China. *Procedia-Social and Behavioral Sciences*, 96, 873-880.

- Hurlin, C. (2003). *Économétrie des Variables Qualitatives - Chapitre 2 - Modèles Multinomiaux - Modèles Logit Multinomiaux Ordonnées et non Ordonnés*. Notes de cours. Tiré de http://www.univ-orleans.fr/deg/masters/ESA/CH/Qualitatif_Chapitre2.pdf
- Hurtubia, R., Gallay, O., & Bierlaire, M. (2010). Attributes of households, locations and real-estate markets for land use modeling. *SustainCity Deliverable*, 2.
- Jarvis, H. (1999a). Housing mobility as a function of household structure: towards a deeper explanation of housing-related disadvantage. *Housing Studies*, 14(4), 491-505.
- Jarvis, H. (1999b). The tangled webs we weave: household strategies to co-ordinate home and work. *Work, Employment & Society*, 13(2), 225-247.
- Kato, H., & Matsumoto, M. (2009). Intra-household interaction in a nuclear family: A utility-maximizing approach. *Transportation Research Part B: Methodological*, 43(2), 191-203.
- Kim, C., & Parent, O. (2016). Modeling individual travel behaviors based on intra-household interactions. *Regional Science and Urban Economics*, 57, 1-11.
- Kim, S. (2007). Changes in the nature of urban spatial structure in the United States, 1890–2000. *Journal of Regional science*, 47(2), 273-287.
- Larousse. (édit.). (2018). *Dictionnaire de français Larousse*.
- Lee, Y., Hickman, M., & Washington, S. (2007). Household type and structure, time-use pattern, and trip-chaining behavior. *Transportation Research Part A: Policy and Practice*, 41(10), 1004-1020.
- Martel Poliquin, É. (2012). *Mieux comprendre les déterminants du choix modal*. (École polytechnique de Montréal).
- Masson, S. (1998). Interactions entre système de transport et système de localisation. De l'héritage des modèles traditionnels à l'apport des modèles interactifs de transport et d'occupation des sols. *Les cahiers scientifiques du transport*(33), 79-108.
- McBride, E. C., Davis, A. W., Lee, J. H., & Goulias, K. G. (2017). Incorporating Land Use into Methods of Synthetic Population Generation and of Transfer of Behavioral Data. *Transportation Research Record: Journal of the Transportation Research Board*(2668), 11-20.
- McDonald, N. C. (2008). Household interactions and children's school travel: the effect of parental work patterns on walking and biking to school. *Journal of transport geography*, 16(5), 324-331.
- McEvedy, C., & Jones, R. (1978). *Atlas of world population history*: Penguin Books Ltd, Harmondsworth, Middlesex, England.
- McQuaid, R. W., & Chen, T. (2012). Commuting times—The role of gender, children and part-time work. *Research in transportation economics*, 34(1), 66-73.
- Miralles-Guasch, C., Melo, M. M., & Marquet, O. (2016). A gender analysis of everyday mobility in urban and rural territories: from challenges to sustainability. *Gender, Place & Culture*, 23(3), 398-417.

- Mohamad, I. B., & Usman, D. (2013). Standardization and its effects on K-means clustering algorithm. *Research Journal of Applied Sciences, Engineering and Technology*, 6(17), 3299-3303.
- Morency, C. (2016). *Cours de Planification des transports CIV 6701* École Polytechnique de Montréal.
- Morency, C., & Trépanier, M. (2013). Enrichissement du processus de modélisation de la demande de transport à l'aide des modèles d'activités. *Montreal, QC: Polytechnique Montreal*.
- MTMDET. (2017a). Concepts et méthodologie des enquêtes origine-destination, Méthodologie générale des enquêtes-ménages. Tiré de <https://www.transports.gouv.qc.ca/fr/ministere/Planification-transports/enquetes-origine-destination/Documents/concept-methodologie.pdf>
- MTMDET. (2017b). *Scénario de prévision de la demande en transport dans la région de Montréal à l'horizon 2036 : prise en compte de nouvelles dimensions tendanciennes* [présentation powerpoint]. Tiré de https://aqtr.com/system/files/file_manager/4_avril_518b_jarouche_fadi_et_timo_dieune_dort.pdf
- Murphy, M. (1991). Modelling households: A synthesis.
- Nock, S. L., & Rossi, P. H. (1978). Household types and social standing. *Soc. F.*, 57, 1325.
- Pelletier, D. (2012). Types de famille et ségrégation résidentielle croisée à Montréal, 2006.
- Polidano, N. (2017). Worldometers : Population mondiale actuelle. Tiré de <http://www.worldometers.info/fr/population-mondiale/#growthrate>
- Pooley, C. G., Horton, D., Scheldeman, G., Tight, M., Jones, T., Chisholm, A., . . . Jopson, A. (2011). Household decision-making for everyday travel: a case study of walking and cycling in Lancaster (UK). *Journal of Transport Geography*, 19(6), 1601-1607.
- Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, 1(1), 81-106. doi:10.1007/bf00116251
- Rakotomalala, R. (2005). Arbres de décision. *Revue Modulad*, 33, 163-187.
- Rakotomalala, R. (2012). Analyse de corrélation: Étude des dépendances-Variables quantitatives. *Document de Cours. Version, 1*. Tiré de https://eric.univ-lyon2.fr/~ricco/cours/cours/Analyse_de_Correlation.pdf
- Reddy, B. S., & Balachandra, P. (2012). Urban mobility: A comparative analysis of megacities of India. *Transport Policy*, 21, 152-164.
- Reverso.). Définition ménage français | dictionnaire français définition synonymes Reverso. Tiré de <http://dictionnaire.reverso.net/francais-definition/m%C3%A9nage>
- Roorda, M. J., Miller, E. J., & Habib, K. M. (2008). Validation of TASHA: A 24-h activity scheduling microsimulation model. *Transportation Research Part A: Policy and Practice*, 42(2), 360-375.
- Rosenbloom, S. (2004). *Understanding women's and men's travel patterns*. Communication présentée à Research on Women's Issues in Transportation: Report of a Conference.

- Roussel, L. (1986). Évolution récente de la structure des ménages dans quelques pays industriels. *Population (french edition)*, 913-934.
- Ruggieri, S. (2017). *Enumerating Distinct Decision Trees*. Communication présentée à Proceedings of the 34th International Conference on Machine Learning, Proceedings of Machine Learning Research. Tiré de <http://proceedings.mlr.press>
- Sandow, E. (2008). Commuting behaviour in sparsely populated areas: evidence from northern Sweden. *Journal of Transport Geography*, 16(1), 14-27.
- Saunier, N. (2016). *Gestion de données en transport CIV8760*. Ecole Polytechnique Montréal.
- Săvoiu, G., & Simăn, I. I. (2017). Population implosion and the model of cavitation applied to demography. *Current Science (00113891)*, 113(1).
- Scheiner, J., & Holz-Rau, C. (2017). Women's complex daily lives: a gendered look at trip chaining and activity pattern entropy in Germany. *Transportation*, 44(1), 117-138.
- Scott, D. M., & Kanaroglou, P. S. (2002). An activity-episode generation model that captures interactions between household heads: development and empirical analysis. *Transportation Research Part B: Methodological*, 36(10), 875-896.
- Secrétariat à l'enquête Origine-Destination. (2013). Enquête OD 2013 - Faits saillants. Tiré de <https://rtm.quebec/Media/Default/pdf/section8/enquete-od-2013-faits-saillants.pdf>
- Séguin, A.-M., & BUISSIERE, Y. (1997). *Household forms and patterns of mobility: The case of the Montréal Metropolitan Area*.
- Shay, E., & Khattak, A. J. (2012). Household travel decision chains: Residential environment, automobile ownership, trips and mode choice. *International Journal of Sustainable Transportation*, 6(2), 88-110.
- Simma, A., & Axhausen, K. (2001). Within-household allocation of travel: Case of upper Austria. *Transportation Research Record: Journal of the Transportation Research Board*(1752), 69-75.
- Solá, A. G. (2016). Constructing work travel inequalities: The role of household gender contracts. *Journal of transport geography*, 53, 32-40.
- Sorrentino, C. (1990). The changing family in international perspective. *Monthly Lab. Rev.*, 113, 41.
- Srinivasan, S., & Bhat, C. R. (2005). Modeling household interactions in daily in-home and out-of-home maintenance activity participation. *Transportation*, 32(5), 523-544.
- Statistique Canada. (2015). Statistical units. Tiré de <https://www.statcan.gc.ca/eng/concepts/units>
- Statistique Canada. (2017). Familles, ménages et état matrimonial : faits saillants du Recensement de 2016. *Le Quotidien*, 11. Tiré de <http://www.statcan.gc.ca/daily-quotidien/170802/dq170802a-fra.pdf>
- Strathman, J. G., Dueker, K. J., & Davis, J. S. (1994). Effects of household structure and selected travel characteristics on trip chaining. *Transportation*, 21(1), 23-45.

- Sweet, M., & Kanaroglou, P. (2016). Gender differences: the role of travel and time use in subjective well-being. *Transportation research part F: traffic psychology and behaviour*, 40, 23-34.
- Syam, A. (2014). *People's travel behaviour in a multicultural society*. Communication présentée à Australasian Transport Research Forum, ATRF 2012 - Proceedings, Perth, WA, Australia.
- Ta, N., Kwan, M. P., Chai, Y., & Liu, Z. (2016). Gendered Space-Time Constraints, Activity Participation and Household Structure: A Case Study Using A GPS-Based Activity Survey in Suburban Beijing, China. *Tijdschrift voor economische en sociale geografie*, 107(5), 505-521.
- Théberge-Barrette, C. (2013). *Modéliser l'impact de la construction résidentielle sur la demande de déplacements*. (École Polytechnique de Montréal).
- Therneau, T. M., & Atkinson, E. J. (1997). An introduction to recursive partitioning using the RPART routines: Technical Report 61. URL <http://www.mayo.edu/hsr/techrpt/61.pdf>.
- Timmermans, H. J., & Zhang, J. (2009). Modeling household activity travel behavior: Examples of state of the art modeling approaches and research agenda. *Transportation Research Part B: Methodological*, 43(2), 187-190.
- Toulemon, L. (2011). Individus, familles, ménages, logements: les compter, les décrire. *Travail, genre et sociétés*(2), 47-66.
- United Nations, D. o. E. a. S. A., Population Division. (2017). *World Population Prospects: The 2017 Revision*.
- Valiquette, F. (2010). *Typologie des chaînes de déplacements et modélisation descriptive des systèmes d'activités des personnes*. (École Polytechnique de Montréal). Tiré de https://publications.polymtl.ca/405/1/2010_FrancoisValiquette.pdf
- Vallin, J., & Meslé, F. (2009). The segmented trend line of highest life expectancies. *Population and Development Review*, 35(1), 159-187.
- Web of science. (2018). Citation report : Household mobility. Tiré de http://apps.webofknowledge.com/CitationReport.do?product=WOS&search_mode=CitationReport&SID=6AxD6kQiVzduo7TxGE9&page=1&cr_pqid=1&viewType=summary&colName=WOS
- Zhang, J., Kuwano, M., Lee, B., & Fujiwara, A. (2009). Modeling household discrete choice behavior incorporating heterogeneous group decision-making mechanisms. *Transportation Research Part B: Methodological*, 43(2), 230-250.

ANNEXES

ANNEXE A : CRÉATION DE L'ID PERSONNE ET DE L'ID MÉNAGE (EXEMPLE AVEC L'ENQUÊTE OD 2013)

```
#obtenir IDPERS

OD13_MTQ_triée$TRY<-
ifelse(shift(OD13_MTQ_triée$NOPER)!=OD13_MTQ_triée$NOPER,TRUE,ifelse(shift(OD
13_MTQ_triée$NOLOG)!=OD13_MTQ_triée$NOLOG, TRUE,FALSE))

OD13_MTQ_triée[1,"TRY"]<-TRUE

OD13_MTQ_triée$IDPERS<-1

for (i in 2:359459){OD13_MTQ_triée[i,"IDPERS"]<-
ifelse(OD13_MTQ_triée[i,"TRY"]==TRUE, 1+OD13_MTQ_triée[i-1,"IDPERS"],
OD13_MTQ_triée[i-1,"IDPERS"])}

#obtenir IDLOG

IDLOG<-data.frame(OD13_MTQ_triée[, "NOLOG"]) #colonne numero de logis

IDLOG$TRUEFALSE<-OD13_MTQ_triée$TLOG

IDLOG<-subset(IDLOG, TRUEFALSE==TRUE)

IDLOG$IDLOG<-(1:66223)

OD13_MTQ_triée<-merge(OD13_MTQ_triée, IDLOG, by="NOLOG")
```

ANNEXE B : CRÉATION DES BD PERSONNES ET MÉNAGES (EXEMPLE AVEC L'ENQUÊTE OD 1998)

```
#creation BD personnes

try1<-aggregate(OD98_MTQ_triée_avecID[72:77],
list(OD98_MTQ_triée_avecID$IDPERS), sum) #motif

try2<-aggregate(OD98_MTQ_triée_avecID["NODEP"],
list(OD98_MTQ_triée_avecID$IDPERS), max) #nodep

try3<-aggregate(OD98_MTQ_triée_avecID[90:103],
list(OD98_MTQ_triée_avecID$IDPERS), sum) #heure, mode et temps

try4<-aggregate(OD98_MTQ_triée_avecID["DISTANCE"],
list(OD98_MTQ_triée_avecID$IDPERS), sum) #distance

try5<-aggregate(OD98_MTQ_triée_avecID["PASSTC"],
list(OD98_MTQ_triée_avecID$IDPERS), max) #PASSTC

library(data.table)

OD98_MTQ_triée_avecID$HORS DOM<-ifelse(shift(OD98_MTQ_triée_avecID$MOTIF)!=11,
ifelse(OD98_MTQ_triée_avecID$HREDEP>shift(OD98_MTQ_triée_avecID$HREDEP),
OD98_MTQ_triée_avecID$HREDEP-shift(OD98_MTQ_triée_avecID$HREDEP),NA),0)
#calcul du temps hors domicile

OD98_MTQ_triée_avecID[1,"HORS DOM"]<-0

try6<-aggregate(OD98_MTQ_triée_avecID["HORS DOM"],
list(OD98_MTQ_triée_avecID$IDPERS), sum) #temps hors domicile

try1<-cbind(try1[, -1], try2[, -1], try3[, -1], try4[, -1], try5[, -1], try6[,
-1])

colnames(try1)[7]<-c("NBDEP")

colnames(try1)[22]<-c("DISTANCE")

colnames(try1)[23]<-c("PASSTC")

colnames(try1)[24]<-c("HORS DOM")

try2<- duplicated(OD98_MTQ_triée_avecID$IDPERS)

OD98_MTQ_triée_avecID<-cbind(OD98_MTQ_triée_avecID, try2)
```

```

OD98_MtQ_triée_avecID_person<-subset(OD98_MtQ_triée_avecID, try2==FALSE)

OD98_MtQ_triée_avecID_person<-OD98_MtQ_triée_avecID_person[-c(72:77, 34,
90:103, 67, 105, 106)] #suppression des colonnes déjà présentes dans try1

OD98_MtQ_triée_avecID_person<-cbind(OD98_MtQ_triée_avecID_person, try1)

OD98_MtQ_triée_avecID_person<-OD98_MtQ_triée_avecID_person[-c(2,34:36,
40:63)]

#nb déplacements sans retour puis sans moins de 5 ans et retour

OD98_MtQ_triée_avecID_person$NBDEPsanretour<-
OD98_MtQ_triée_avecID_person$NBDEP - OD98_MtQ_triée_avecID_person$RETOUR

OD98_MtQ_triée_avecID_person$NBDEPsansless5niretour<-
ifelse(OD98_MtQ_triée_avecID_person$AGE <5, 0,
OD98_MtQ_triée_avecID_person$NBDEPsanretour)

OD98_MtQ_triée_avecID_person$NBDEPsansless5niretourten<-
ifelse(OD98_MtQ_triée_avecID_person$NBDEPsansless5niretour>=10,
10,OD98_MtQ_triée_avecID_person$NBDEPsansless5niretour)

OD98_MtQ_triée_avecID_person$NBDEPsansless5niretourfive<-
ifelse(OD98_MtQ_triée_avecID_person$NBDEPsansless5niretour>=5,
5,OD98_MtQ_triée_avecID_person$NBDEPsansless5niretour)

OD98_MtQ_triée_avecID_person$NBDEPsansless5niretoursix<-
ifelse(OD98_MtQ_triée_avecID_person$NBDEPsansless5niretour>=6,
6,OD98_MtQ_triée_avecID_person$NBDEPsansless5niretour)

#obtenir NB et TAILLE chaînes dans BD personnes

try1<-OD98_MtQ_triée_avecID[c("IDPERS", "NOLOG", "NOPER")]

colnames(chaines_1998_fp)[1]<-c("NOLOG")

colnames(chaines_1998_fp)[2]<-c("NOPER")

try<-merge(chaines_1998_fp, try1, by=c("NOLOG", "NOPER"), all=FALSE)

try<-aggregate(try, list(try$CH_IPERE), max)

try1<-aggregate(try["CH_NO"], list(try$IDPERS), max) #nb chaîne

try2<-aggregate(try["NB_DEPL"], list(try$IDPERS), sum) #somme nb depl

```

```

try3<-aggregate(try[20:22], list(try$IDPERS), sum) #chaînes simples,
contrainte, ouvertes

try4<-aggregate(try["IDPERS"], list(try$IDPERS), max) #IDPERS

chaines_avec_id<-cbind(try4[,-1], try1[,-1], try2[,-1], try3[,-1])

colnames(chaines_avec_id)[1]<-c("IDPERS")

colnames(chaines_avec_id)[2]<-c("CH_NO")

colnames(chaines_avec_id)[3]<-c("NB_DEPL")

OD98_MTQ_triée_avecID_person<-merge(OD98_MTQ_triée_avecID_person,
chaines_avec_id, by="IDPERS", all=TRUE)

colnames(OD98_MTQ_triée_avecID_person)[86]<-c("CHAINE_SIMPLE")

colnames(OD98_MTQ_triée_avecID_person)[87]<-c("CHAINE_CONTRAINTTE")

colnames(OD98_MTQ_triée_avecID_person)[88]<-c("CHAINE_OUVERTE")


#calculs d'attributs de la personne

OD98_MTQ_triée_avecID_person$ISZEROTRIPPER<-
ifelse(OD98_MTQ_triée_avecID_person$NBDEP==0, 1, 0)

OD98_MTQ_triée_avecID_person$DISTMOY<-
OD98_MTQ_triée_avecID_person$DISTANCE/OD98_MTQ_triée_avecID_person$NBDEP

OD98_MTQ_triée_avecID_person$PARTTA<-
OD98_MTQ_triée_avecID_person$TA/OD98_MTQ_triée_avecID_person$NBDEP

OD98_MTQ_triée_avecID_person$PARTTC<-
OD98_MTQ_triée_avecID_person$TC/OD98_MTQ_triée_avecID_person$NBDEP

OD98_MTQ_triée_avecID_person$PARTAUTO<-
OD98_MTQ_triée_avecID_person$auto/OD98_MTQ_triée_avecID_person$NBDEP

OD98_MTQ_triée_avecID_person$DsurND<-
ifelse(OD98_MTQ_triée_avecID_person$DOUX!=0,
ifelse(OD98_MTQ_triée_avecID_person$NONDOUX!=0,
OD98_MTQ_triée_avecID_person$DOUX/OD98_MTQ_triée_avecID_person$NONDOUX, NA),
NA)

OD98_MTQ_triée_avecID_person$TAsurAUTO<-
ifelse(OD98_MTQ_triée_avecID_person$TA!=0,

```

```

ifelse(OD98_MtQ_triée_avecID_person$auto!=0,
OD98_MtQ_triée_avecID_person$TA/OD98_MtQ_triée_avecID_person$auto, NA), NA)

OD98_MtQ_triée_avecID_person$TCsurAUTO<-
ifelse(OD98_MtQ_triée_avecID_person$TC!=0,
ifelse(OD98_MtQ_triée_avecID_person$auto!=0,
OD98_MtQ_triée_avecID_person$TC/OD98_MtQ_triée_avecID_person$auto, NA), NA)

OD98_MtQ_triée_avecID_person$PARTPOINTEAM<-OD98_MtQ_triée_avecID_person$hre2/
OD98_MtQ_triée_avecID_person$NBDEP

OD98_MtQ_triée_avecID_person$PARTPOINTEPM<-OD98_MtQ_triée_avecID_person$hre2/
OD98_MtQ_triée_avecID_person$NBDEP

OD98_MtQ_triée_avecID_person$CHAINE_COMPLEXE<-
ifelse(OD98_MtQ_triée_avecID_person$CHAINE_SIMPLE==0,
OD98_MtQ_triée_avecID_person$CH_NO-
OD98_MtQ_triée_avecID_person$CHAINE_SIMPLE,0)

OD98_MtQ_triée_avecID_person$INF_6ANS<-
ifelse(OD98_MtQ_triée_avecID_person$AGE<6, 1, 0)

OD98_MtQ_triée_avecID_person$INF_12ANS<-
ifelse(OD98_MtQ_triée_avecID_person$AGE<12, 1, 0)

OD98_MtQ_triée_avecID_person$ENTRE6_11ANS<-
ifelse(OD98_MtQ_triée_avecID_person$AGE<12,
ifelse(OD98_MtQ_triée_avecID_person$AGE>5, 1, 0),0)

OD98_MtQ_triée_avecID_person$ENTRE12_16ANS<-
ifelse(OD98_MtQ_triée_avecID_person$AGE<17,
ifelse(OD98_MtQ_triée_avecID_person$AGE>11, 1, 0),0)

OD98_MtQ_triée_avecID_person$ENTRE76_84ANS<-
ifelse(OD98_MtQ_triée_avecID_person$AGE<85,
ifelse(OD98_MtQ_triée_avecID_person$AGE>76, 1, 0),0)

OD98_MtQ_triée_avecID_person$SUP_84ANS<-
ifelse(OD98_MtQ_triée_avecID_person$AGE>84, 1, 0)

#creation BD ménages

OD98_MtQ_triée_avecID_person$FACPERA<-0

```

```

try0<-aggregate(OD98_MTQ_triée_avecID_person[c("IDLOG")],
list(OD98_MTQ_triée_avecID_person$IDLOG), max) #IDLOG

try1<-aggregate(OD98_MTQ_triée_avecID_person[c(38:53)],
list(OD98_MTQ_triée_avecID_person$IDLOG), sum) #sexe et occupation des
personnes

try2<-aggregate(OD98_MTQ_triée_avecID_person[c(55:78)],
list(OD98_MTQ_triée_avecID_person$IDLOG), sum) #motif, heure et mode, temps,
distance

try3<-aggregate(OD98_MTQ_triée_avecID_person[c(79:89)],
list(OD98_MTQ_triée_avecID_person$IDLOG), sum) #passtc, sans retour et sans
retour ni less5 et chaînes et zero tripper

try4<-aggregate(OD98_MTQ_triée_avecID_person[c("CHAINE_COMPLEXE")],
list(OD98_MTQ_triée_avecID_person$IDLOG), sum) #chaîne complexe

try5<-aggregate(OD98_MTQ_triée_avecID_person[c(24:26)],
list(OD98_MTQ_triée_avecID_person$IDLOG), sum) #percond, facper

try6<-aggregate(OD98_MTQ_triée_avecID_person[c(100:105)],
list(OD98_MTQ_triée_avecID_person$IDLOG), sum) #catés age

try1<-cbind( try0[,-1], try1[,-1], try2[,-1], try3[,-1], try4[,-1], try5[,-
1], try6[,-1])

colnames(try1)[1]<-c("IDLOG")

try2<- duplicated(OD98_MTQ_triée_avecID_person$IDLOG)

OD98_MTQ_triée_avecID_person<-cbind(OD98_MTQ_triée_avecID_person, try2)

OD98_MTQ_triée_avecID_household<-subset(OD98_MTQ_triée_avecID_person,
try2==FALSE)

OD98_MTQ_triée_avecID_household<-OD98_MTQ_triée_avecID_household[-c(38:53,
55:78, 79:89, 19:26, 99, 100:105)]

OD98_MTQ_triée_avecID_household<-OD98_MTQ_triée_avecID_household[-c(1,16:28,
31:40)]

OD98_MTQ_triée_avecID_household<-merge(OD98_MTQ_triée_avecID_household, try1,
by="IDLOG", all=TRUE)

#colnames(OD98_MTQ_triée_avecID_household)[66]<-c("NBZEROTRIPPER")

#colnames(OD98_MTQ_triée_avecID_household)[63]<-c("CHAINE_SIMPLE")

```

```

#colnames(OD98_MTQ_triée_avecID_household)[64]<-c("CHAINE_CONTRAINTE")

#colnames(OD98_MTQ_triée_avecID_household)[65]<-c("CHAINE_OUVERTE")

colnames(OD98_MTQ_triée_avecID_household)[68]<-c("CHAINE_COMPLEXE")


#calculs d'attributs du ménage

OD98_MTQ_triée_avecID_household$DISTMOY<-
OD98_MTQ_triée_avecID_household$DISTANCE/OD98_MTQ_triée_avecID_household$NBDE
P

OD98_MTQ_triée_avecID_household$PARTTA<-
OD98_MTQ_triée_avecID_household$TA/OD98_MTQ_triée_avecID_household$NBDEP

OD98_MTQ_triée_avecID_household$PARTTC<-
OD98_MTQ_triée_avecID_household$TC/OD98_MTQ_triée_avecID_household$NBDEP

OD98_MTQ_triée_avecID_household$PARTAUTO<-
OD98_MTQ_triée_avecID_household$auto/OD98_MTQ_triée_avecID_household$NBDEP

OD98_MTQ_triée_avecID_household$DsurND<-
ifelse(OD98_MTQ_triée_avecID_household$DOUX!=0,
ifelse(OD98_MTQ_triée_avecID_household$NONDOUX!=0,
OD98_MTQ_triée_avecID_household$DOUX/OD98_MTQ_triée_avecID_household$NONDOUX,
NA), NA)

OD98_MTQ_triée_avecID_household$TAsurAUTO<-
ifelse(OD98_MTQ_triée_avecID_household$TA!=0,
ifelse(OD98_MTQ_triée_avecID_household$auto!=0,
OD98_MTQ_triée_avecID_household$TA/OD98_MTQ_triée_avecID_household$auto, NA),
NA)

OD98_MTQ_triée_avecID_household$TCsurAUTO<-
ifelse(OD98_MTQ_triée_avecID_household$TC!=0,
ifelse(OD98_MTQ_triée_avecID_household$auto!=0,
OD98_MTQ_triée_avecID_household$TC/OD98_MTQ_triée_avecID_household$auto, NA),
NA)

OD98_MTQ_triée_avecID_household$PARTPOINTEAM<-
OD98_MTQ_triée_avecID_household$hre2/ OD98_MTQ_triée_avecID_household$NBDEP

OD98_MTQ_triée_avecID_household$PARTPOINTEPM<-
OD98_MTQ_triée_avecID_household$hre2/ OD98_MTQ_triée_avecID_household$NBDEP

```

```
OD98_MTQ_triée_avecID_household$ACCESAUTO<-
OD98_MTQ_triée_avecID_household$NBVEH/OD98_MTQ_triée_avecID_household$NBPERIN
F16
```

```
#retrouver les âges de chaque personne du ménage pour trouver l'âge moyen etc
```

```
essaie98<-aggregate(OD98_MTQ_triée_avecID_person["AGE"],
list(OD98_MTQ_triée_avecID_person$IDLOG), FUN=as.list)

for(j in 1:14){
  for(i in 1:65227){
    essaie98[i,j+2] <-
ifelse(length(essaie98$AGE[[i]])>=j,essaie98$AGE[[i]][[j]],NA)
  }
}

colnames(essaie98)[3]<-c("pers1")
colnames(essaie98)[4]<-c("pers2")
colnames(essaie98)[5]<-c("pers3")
colnames(essaie98)[6]<-c("pers4")
colnames(essaie98)[7]<-c("pers5")
colnames(essaie98)[8]<-c("pers6")
colnames(essaie98)[9]<-c("pers7")
colnames(essaie98)[10]<-c("pers8")
colnames(essaie98)[11]<-c("pers9")
colnames(essaie98)[12]<-c("pers10")
colnames(essaie98)[13]<-c("pers11")
colnames(essaie98)[14]<-c("pers12")
colnames(essaie98)[15]<-c("pers13")
colnames(essaie98)[16]<-c("pers14")

library(matrixStats)

essaie98$AGEMEAN<-rowMeans(essaie98[, 3:16], na.rm=TRUE)
```

```
essaie98$AGESD<-rowSds(as.matrix(essaie98[, 3:16]), na.rm=TRUE)
essaie98$AGEMIN<-rowMins(as.matrix(essaie98[, 3:16]), na.rm=TRUE)
essaie98$AGEMAX<-rowMaxs(as.matrix(essaie98[, 3:16]), na.rm=TRUE)
essaie98$AGEAMPMAX<-essaie98$AGEMAX-essaie98$AGEMIN
colnames(essaie98)[1]<-c("IDLOG")
NBPER<-OD98_MTQ_triée_avecID_household[c(1,8)]
essaie98<-merge(essaie98, NBPER, by=c("IDLOG"), all=FALSE)
ages_BD98<-ages_BD98[-c(1)]
OD98_MTQ_triée_avecID_household<-merge(OD98_MTQ_triée_avecID_household,
ages_BD98, by="IDLOG", all=TRUE)
```

ANNEXE C : CONSTRUCTION D'UN ARBRE ET DE SA MATRICE DE CONFUSION (EXEMPLE AVEC 3 DÉPLACEMENTS POUR 2013)

```
#3dep

arbre_threedep13<-rpart(threedep ~ SEXE + AGE + REG_DOM + NB_HOMMES_MENAGE +
NB_FEMMES_MENAGE + NB_PERSONNES_MENAGE + pres_INF_12ANS + pres_ENTRE12_16ANS
+ pres_ENTRE76_84ANS + pres_SUP_84ANS + AGE_AMP_MAX + AGE_MEAN + ACCES_AUTO,
data=OD13_MTO_arbre_sans04ans_v2, method= "class", control =
rpart.control(minsplit=70, cp=0.00001))

print(arbre_threedep13)

pdf("arbre_threedep13.pdf")

prp(arbre_threedep13, gap=0, space=0, uniform=T, type=4, faclen=0,
compress=F,extra=108, branch=1, yesno=FALSE, border.col=0, xsep="/")

dev.off()

plotcp(arbre_threedep13)

printcp(arbre_threedep13)

plot(predict(arbre_threedep13, type="class"),
OD13_MTO_arbre_sans04ans_v2$threedep)

table(predict(arbre_threedep13, type="class"),
OD13_MTO_arbre_sans04ans_v2$threedep)

summary(arbre_threedep13)

arbre_3dep_13_13avr<-arbre_threedep13

table(predict(arbre_3dep_13_13avr, type="class"),
na.omit(OD13_MTO_arbre_sans04ans_v2$threedep))

library(ggplot2)

library(textir)

input.matrix <- data.matrix(table(predict(arbre_3dep_13_13avr, type="class"),
OD13_MTO_arbre_sans04ans_v2$threedep))

colnames(input.matrix) = c("OUI", "NON")

rownames(input.matrix) = colnames(input.matrix)

confusion <- as.data.frame(as.table(input.matrix))
```

```
pdf("Matrice de confusion arbre 3 dep 2013.pdf", paper='A4r', width=1500)

plot <- ggplot(confusion)

plot + geom_tile(aes(x=Var1, y=Var2, fill=Freq)) +
scale_x_discrete(name="Prédit") + scale_y_discrete(name="Observé") +
labs(fill="Fréquence\nnormalisée")+ theme(legend.position = "right",
legend.text = element_text(size=18), legend.title = element_text(size=20),
axis.title.x = element_text(size=20), axis.title.y = element_text(size=20),
axis.text.x = element_text(size=15), axis.text.y = element_text(size=15))

dev.off()
```

ANNEXE D : CONSTRUCTION DE LA TYPOLOGIE (EXEMPLE AVEC LA BASE DE DONNÉES COMPRENANT LES QUATRE ANNÉES)

```

cluster<-kmeans(OD98a13_MTQ_cluster[,c(5, 8:9, 11, 14)], 8)

print(cluster)

cluster_menage_98a13<-cluster

OD98a13_MTQ_cluster$cluster<-cluster$cluster

OD98a13_MTQ_cluster$cluster_comparaison<-
ifelse(OD98a13_MTQ_cluster$cluster==7, 3,
ifelse(OD98a13_MTQ_cluster$cluster==1, 8,
ifelse(OD98a13_MTQ_cluster$cluster==5, 1,
ifelse(OD98a13_MTQ_cluster$cluster==3, 6,
ifelse(OD98a13_MTQ_cluster$cluster==8, 7,
ifelse(OD98a13_MTQ_cluster$cluster==6, 4,
ifelse(OD98a13_MTQ_cluster$cluster==2, 5,
ifelse(OD98a13_MTQ_cluster$cluster==4, 2, NA)))))))))

write.csv(OD98a13_MTQ_cluster, file = "OD98a13_MTQ_cluster.csv",
row.names=TRUE)

OD98a13_MTQ_cluster$AGE_MEAN_cate <- cut(OD98a13_MTQ_cluster$AGE_MEAN, breaks
= c(0, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90,
95, 100, Inf), right = FALSE)

OD98a13_MTQ_cluster$AGE_AMP_MAX_cate <- cut(OD98a13_MTQ_cluster$AGE_AMP_MAX,
breaks = c(0, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80,
85, 90, 95, 100, Inf), right = FALSE)

OD98a13_MTQ_cluster$PART_DEP_AUTO_AUTOPASSAGER_cate <-
cut(OD98a13_MTQ_cluster$PART_DEP_AUTO_AUTOPASSAGER, breaks = c(0, 0.2, 0.4,
0.6, 0.8, 1, 2, Inf), right = FALSE)

OD98a13_MTQ_cluster$PART_DEP_POINTE_AM_cate <-
cut(OD98a13_MTQ_cluster$PART_DEP_POINTE_AM, breaks = c(0, 0.2, 0.4, 0.6, 0.8,
1, 2, Inf), right = FALSE)

#Affichage des groupes, de leur taille et de leurs moyennes

```

```

pdf("Clusters suivant caractéristiques du ménage 1998 à 2013.pdf",
paper='A4r', width=1500)

ggplot(OD98a13_MTQ_cluster, aes(AGE_MEAN, AGE_AMP_MAX, color =
OD98a13_MTQ_cluster$cluster_comparaison)) +

  geom_point(aes(colour = factor(cluster_comparaison))) +

  labs(y = "Amplitude maximale des âges dans le ménage", x="Âge moyen du
ménage") +

  scale_color_brewer(name="Cluster \n", palette="Accent")

dev.off()

stargazer(cluster_menage_98a13$size, title="taille des clusters",
colnames=TRUE, style="commadefault", type= "html",
out="size_cluster_men_98a13.html")

stargazer(cluster_menage_98a13$centers, colnames=TRUE, rownames=TRUE,
style="commadefault", type= "html", out="centers_cluster_men_98a13.html")

#comparaison des comportements de déplacement

pdf("Chaîne complexe par cluster 1998 à 2013.pdf", paper='A4r', width=1500)

ggplot(OD98a13_MTQ_cluster_sanszerotripper,
aes(x=AU_MOINS_UNE_CHAINE_COMPLEXE, group=cluster_comparaison)) +

  geom_bar(aes(y = ..prop.., fill = factor(..x..)), stat="count") +

  geom_text_repel(aes( label = scales ::percent(..prop..),
                        y= ..prop.. ), stat= "count", vjust = -.5, hjust= 0.1,
size=4)+

  labs(y = "Pourcentage de la population par groupe", x="Types de ménages") +

  facet_grid(~cluster_comparaison) +

  scale_y_continuous(labels = scales::percent)+

  theme(axis.text.x = element_blank(),

        axis.ticks.x = element_blank(),

        legend.position = "right", legend.text = element_text(size=18),
legend.title = element_text(size=20), axis.title.y = element_text(size=20),
axis.title.x = element_text(size=20), axis.text.y = element_text(size=15))+

```

```

scale_fill_discrete(name="Faire au moins \nune chaîne \ncomplexe \n",
                    breaks=c("1", "2", "3"),
                    labels=c("NON", "OUI", "Donnée indisponible"))

dev.off()

pdf("Part pointe am par cluster 1998 à 2013.pdf", paper='A4r', width=1500)

ggplot(OD98a13_MTQ_cluster_sanszerotripper, aes(x=PART_DEP_POINTE_AM_cate,
group=cluster_comparaison)) +

  geom_bar(aes(y = ..prop.., fill = factor(..x..)), stat="count") +

  geom_text_repel(aes( label = scales::percent(..prop..),
                      y= ..prop.. ), stat= "count", vjust = -.5, hjust= 0.1,
size=4)+

  labs(y = "Pourcentage de la population par groupe", x="Types de ménages") +

  facet_grid(~cluster_comparaison) +

  scale_y_continuous(labels = scales::percent)+

  theme(axis.text.x = element_blank(),

        axis.ticks.x = element_blank(),

        legend.position = "right", legend.text = element_text(size=18),
legend.title = element_text(size=20), axis.title.y = element_text(size=20),
axis.title.x = element_text(size=20), axis.text.y = element_text(size=15))+

  scale_fill_discrete(name="Part de \ndéplacements \nen pointe AM \n",
                    breaks=c("1", "2", "3", "4", "5", "6"),
                    labels=c("[0, 0.2)", "[0.2, 0.4)", "[0.4, 0.6)", "[0.6,
0.8)", "[0.8, 1)", "1"))

dev.off()

```

```

pdf("Part auto par cluster 1998 à 2013.pdf", paper='A4r', width=1500)

ggplot(OD98a13_MTQ_cluster_sanszerotripper,
aes(x=PART_DEP_AUTO_AUTOPASSAGER_cate, group=cluster_comparaison)) +

  geom_bar(aes(y = ..prop.., fill = factor(..x..)), stat="count") +

```

```

geom_text_repel(aes( label = scales::percent(..prop..),
                    y= ..prop.. ), stat= "count", vjust = -.5, hjust= 0.1,
size=4)+

labs(y = "Pourcentage de la population par groupe", x="Types de ménages") +
facet_grid(~cluster_comparaison) +
scale_y_continuous(labels = scales::percent)+
theme(axis.text.x = element_blank(),
      axis.ticks.x = element_blank(),
      legend.position = "right", legend.text = element_text(size=18),
legend.title = element_text(size=20), axis.title.y = element_text(size=20),
axis.title.x = element_text(size=20), axis.text.y = element_text(size=15))+
  scale_fill_discrete(name="Part de \ndéplacements \nen auto \n(conducteur
\nou passager) \n",
                      breaks=c("1", "2", "3", "4", "5", "6"),
                      labels=c("[0, 0.2)", "[0.2, 0.4)", "[0.4, 0.6)", "[0.6,
0.8)", "[0.8, 1)", "1"))
dev.off()

library(ggrepel)

pdf("nombre dep par cluster 1998 à 2013.pdf", paper='A4r', width=1500)

ggplot(OD98a13_MTQ_cluster, aes(x=NB_DEP, group=cluster_comparaison)) +
  geom_bar(aes(y = ..prop.., fill = factor(..x..)), stat="count") +

  geom_text_repel(aes( label = scales::percent(..prop..),
                    y= ..prop.. ), stat= "count", vjust = -.5, hjust= 0.1,
size=4)+

labs(y = "Pourcentage de la population par groupe", x="Types de ménages") +
facet_grid(~cluster_comparaison) +
scale_y_continuous(labels = scales::percent)+
theme(axis.text.x = element_blank(),
      axis.ticks.x = element_blank(),

```


ANNEXE E : CONSTRUCTION DU MODÈLE LOGIT MULTINOMIAL ET DE LA MATRICE DE CONFUSION

```

logit98a13 <- multinom(cluster_comparaison ~ HOMME + FEMME + AGE + REG1 +
REG2 + REG3 + REG4 + REG5 + REG6 + REG7 + REG8 + NB_ANNEE_SOURCE, data =
OD98a13_MTQ_logit)

logit98a13step <- step(logit98a13)

summary(logit98a13)

summary(logit98a13step)

logit98a13_oddsratio<-odds.ratio(logit98a13step)

probabilities_logit98a13<-predict(logit98a13step, newdata =
OD98a13_MTQ_logit, type ="probs")

OD98a13_MTQ_logit<-cbind(OD98a13_MTQ_logit, probabilities_logit98a13)


#10 itérations pour la matrice de confusion

OD98a13_MTQ_logit_bis<-OD98a13_MTQ_logit_bis[-c(1:2, 5:6)]

OD98a13_MTQ_logit_bis$random1<-runif(646251, min = 0, max = 1)
OD98a13_MTQ_logit_bis$random2<-runif(646251, min = 0, max = 1)
OD98a13_MTQ_logit_bis$random3<-runif(646251, min = 0, max = 1)
OD98a13_MTQ_logit_bis$random4<-runif(646251, min = 0, max = 1)
OD98a13_MTQ_logit_bis$random5<-runif(646251, min = 0, max = 1)
OD98a13_MTQ_logit_bis$random6<-runif(646251, min = 0, max = 1)
OD98a13_MTQ_logit_bis$random7<-runif(646251, min = 0, max = 1)
OD98a13_MTQ_logit_bis$random8<-runif(646251, min = 0, max = 1)
OD98a13_MTQ_logit_bis$random9<-runif(646251, min = 0, max = 1)
OD98a13_MTQ_logit_bis$random10<-runif(646251, min = 0, max = 1)


colnames(OD98a13_MTQ_logit_bis)[60:67]<-c("un", "deux", "trois", "quatre",
"cinq", "six", "sept", "huit")

OD98a13_MTQ_logit_bis$cumul1<-OD98a13_MTQ_logit_bis$un

```

```

OD98a13_MTQ_logit_bis$cumul2<-OD98a13_MTQ_logit_bis$un +
OD98a13_MTQ_logit_bis$deux

OD98a13_MTQ_logit_bis$cumul3<-OD98a13_MTQ_logit_bis$un +
OD98a13_MTQ_logit_bis$deux + OD98a13_MTQ_logit_bis$trois

OD98a13_MTQ_logit_bis$cumul4<-OD98a13_MTQ_logit_bis$un +
OD98a13_MTQ_logit_bis$deux + OD98a13_MTQ_logit_bis$trois +
OD98a13_MTQ_logit_bis$quatre

OD98a13_MTQ_logit_bis$cumul5<-OD98a13_MTQ_logit_bis$un +
OD98a13_MTQ_logit_bis$deux + OD98a13_MTQ_logit_bis$trois +
OD98a13_MTQ_logit_bis$quatre + OD98a13_MTQ_logit_bis$cinq

OD98a13_MTQ_logit_bis$cumul6<-OD98a13_MTQ_logit_bis$un +
OD98a13_MTQ_logit_bis$deux + OD98a13_MTQ_logit_bis$trois +
OD98a13_MTQ_logit_bis$quatre + OD98a13_MTQ_logit_bis$cinq +
OD98a13_MTQ_logit_bis$six

OD98a13_MTQ_logit_bis$cumul7<-OD98a13_MTQ_logit_bis$un +
OD98a13_MTQ_logit_bis$deux + OD98a13_MTQ_logit_bis$trois +
OD98a13_MTQ_logit_bis$quatre + OD98a13_MTQ_logit_bis$cinq +
OD98a13_MTQ_logit_bis$six + OD98a13_MTQ_logit_bis$sept

OD98a13_MTQ_logit_bis$cumul8<-OD98a13_MTQ_logit_bis$un +
OD98a13_MTQ_logit_bis$deux + OD98a13_MTQ_logit_bis$trois +
OD98a13_MTQ_logit_bis$quatre + OD98a13_MTQ_logit_bis$cinq +
OD98a13_MTQ_logit_bis$six + OD98a13_MTQ_logit_bis$sept +
OD98a13_MTQ_logit_bis$huit

OD98a13_MTQ_logit_bis$predict1<-
ifelse(OD98a13_MTQ_logit_bis$random1<OD98a13_MTQ_logit_bis$cumul1, 1,
ifelse(OD98a13_MTQ_logit_bis$random1<OD98a13_MTQ_logit_bis$cumul2, 2,
ifelse(OD98a13_MTQ_logit_bis$random1<OD98a13_MTQ_logit_bis$cumul3, 3,
ifelse(OD98a13_MTQ_logit_bis$random1<OD98a13_MTQ_logit_bis$cumul4, 4,
ifelse(OD98a13_MTQ_logit_bis$random1<OD98a13_MTQ_logit_bis$cumul5, 5,
ifelse(OD98a13_MTQ_logit_bis$random1<OD98a13_MTQ_logit_bis$cumul6, 6,
ifelse(OD98a13_MTQ_logit_bis$random1<OD98a13_MTQ_logit_bis$cumul7, 7,
ifelse(OD98a13_MTQ_logit_bis$random1<OD98a13_MTQ_logit_bis$cumul8, 8))))))

OD98a13_MTQ_logit_bis$predict2<-
ifelse(OD98a13_MTQ_logit_bis$random2<OD98a13_MTQ_logit_bis$cumul1, 1,

```

```

ifelse(OD98a13_MTQ_logit_bis$random2<OD98a13_MTQ_logit_bis$cumul2, 2,
ifelse(OD98a13_MTQ_logit_bis$random2<OD98a13_MTQ_logit_bis$cumul3, 3,
ifelse(OD98a13_MTQ_logit_bis$random2<OD98a13_MTQ_logit_bis$cumul4, 4,
ifelse(OD98a13_MTQ_logit_bis$random2<OD98a13_MTQ_logit_bis$cumul5, 5,
ifelse(OD98a13_MTQ_logit_bis$random2<OD98a13_MTQ_logit_bis$cumul6, 6,
ifelse(OD98a13_MTQ_logit_bis$random2<OD98a13_MTQ_logit_bis$cumul7, 7,
ifelse(OD98a13_MTQ_logit_bis$random2<OD98a13_MTQ_logit_bis$cumul8, 8)))))))))

```

```

ifelse(OD98a13_MTQ_logit_bis$random6<OD98a13_MTQ_logit_bis$cumul3, 3,
ifelse(OD98a13_MTQ_logit_bis$random6<OD98a13_MTQ_logit_bis$cumul4, 4,
ifelse(OD98a13_MTQ_logit_bis$random6<OD98a13_MTQ_logit_bis$cumul5, 5,
ifelse(OD98a13_MTQ_logit_bis$random6<OD98a13_MTQ_logit_bis$cumul6, 6,
ifelse(OD98a13_MTQ_logit_bis$random6<OD98a13_MTQ_logit_bis$cumul7, 7,
ifelse(OD98a13_MTQ_logit_bis$random6<OD98a13_MTQ_logit_bis$cumul8, 8))))))

```

```

ifelse(OD98a13_MTQ_logit_bis$random10<OD98a13_MTQ_logit_bis$cumul4, 4,
ifelse(OD98a13_MTQ_logit_bis$random10<OD98a13_MTQ_logit_bis$cumul5, 5,
ifelse(OD98a13_MTQ_logit_bis$random10<OD98a13_MTQ_logit_bis$cumul6, 6,
ifelse(OD98a13_MTQ_logit_bis$random10<OD98a13_MTQ_logit_bis$cumul7, 7,
ifelse(OD98a13_MTQ_logit_bis$random10<OD98a13_MTQ_logit_bis$cumul8, 8)))))))))

```

```
library(ggplot2)
```

```
logit98a13_conf_matrix1<-table(OD98a13_MTQ_logit_bis$predict1,
OD98a13_MTQ_logit_bis$cluster_comparaison)
```

```
input.matrix <- data.matrix(logit98a13_conf_matrix1)
```

```
colnames(input.matrix) = c("1", "2", "3", "4", "5", "6", "7", "8")
```

```
rownames(input.matrix) = colnames(input.matrix)
```

```
confusion1 <- as.data.frame(as.table(input.matrix))
```

```
pdf("Matrice de confusion modèle d'affectation 1998 à 2013 1.pdf",
paper='A4r', width=1500)
```

```
plot <- ggplot(confusion1)
```

```
plot + geom_tile(aes(x=Var1, y=Var2, fill=Freq)) +
scale_x_discrete(name="Type de ménage prédit") + scale_y_discrete(name="Type
de ménage observé") + scale_fill_gradient(breaks=seq(from=-.5, to=4, by=.2))
+ labs(fill="Fréquence\nnormalisée") + theme(legend.position = "right",
legend.text = element_text(size=18), legend.title = element_text(size=20),
axis.title.x = element_text(size=20), axis.title.y = element_text(size=20),
axis.text.x = element_text(size=15), axis.text.y = element_text(size=15))
```

```
dev.off()
```

```
logit98a13_conf_matrix2<-table(OD98a13_MTQ_logit_bis$predict2,
OD98a13_MTQ_logit_bis$cluster_comparaison)
```

```
input.matrix <- data.matrix(logit98a13_conf_matrix2)
```

```
colnames(input.matrix) = c("1", "2", "3", "4", "5", "6", "7", "8")
```

```
rownames(input.matrix) = colnames(input.matrix)
```

```
confusion2 <- as.data.frame(as.table(input.matrix))
```

```

pdf("Matrice de confusion modèle d'affectation 1998 à 2013 2.pdf",
paper='A4r', width=1500)

plot <- ggplot(confusion2)

plot + geom_tile(aes(x=Var1, y=Var2, fill=Freq)) +
scale_x_discrete(name="Type de ménage prédit") + scale_y_discrete(name="Type
de ménage observé") + scale_fill_gradient(breaks=seq(from=-.5, to=4, by=.2))
+ labs(fill="Fréquence\nnormalisée") + theme(legend.text =
element_text(size=18), legend.title = element_text(size=20), axis.title.x =
element_text(size=20), axis.title.y = element_text(size=20), axis.text.x =
element_text(size=15), axis.text.y = element_text(size=15))

dev.off()

logit98a13_conf_matrix3<-table(OD98a13_MTQ_logit_bis$predict3,
OD98a13_MTQ_logit_bis$cluster_comparaison)

input.matrix <- data.matrix(logit98a13_conf_matrix3)

colnames(input.matrix) = c("1", "2", "3", "4", "5", "6", "7", "8")

rownames(input.matrix) = colnames(input.matrix)

confusion3 <- as.data.frame(as.table(input.matrix))

pdf("Matrice de confusion modèle d'affectation 1998 à 2013 3.pdf",
paper='A4r', width=1500)

plot <- ggplot(confusion3)

plot + geom_tile(aes(x=Var1, y=Var2, fill=Freq)) +
scale_x_discrete(name="Type de ménage prédit") + scale_y_discrete(name="Type
de ménage observé") + scale_fill_gradient(breaks=seq(from=-.5, to=4, by=.2))
+ labs(fill="Fréquence\nnormalisée") + theme(legend.text =
element_text(size=18), legend.title = element_text(size=20), axis.title.x =
element_text(size=20), axis.title.y = element_text(size=20), axis.text.x =
element_text(size=15), axis.text.y = element_text(size=15))

dev.off()

logit98a13_conf_matrix4<-table(OD98a13_MTQ_logit_bis$predict4,
OD98a13_MTQ_logit_bis$cluster_comparaison)

```

```

input.matrix <- data.matrix(logit98a13_conf_matrix4)

colnames(input.matrix) = c("1", "2", "3", "4", "5", "6", "7", "8")

rownames(input.matrix) = colnames(input.matrix)

confusion4 <- as.data.frame(as.table(input.matrix))

pdf("Matrice de confusion modèle d'affectation 1998 à 2013 4.pdf",
    paper='A4r', width=1500)

plot <- ggplot(confusion4)

plot + geom_tile(aes(x=Var1, y=Var2, fill=Freq)) +
  scale_x_discrete(name="Type de ménage prédit") + scale_y_discrete(name="Type
de ménage observé") + scale_fill_gradient(breaks=seq(from=-.5, to=4, by=.2))
+ labs(fill="Fréquence\nnormalisée") + theme(legend.text =
  element_text(size=18), legend.title = element_text(size=20), axis.title.x =
  element_text(size=20), axis.title.y = element_text(size=20), axis.text.x =
  element_text(size=15), axis.text.y = element_text(size=15))

dev.off()

logit98a13_conf_matrix5<-table(OD98a13_MTQ_logit_bis$predict5,
OD98a13_MTQ_logit_bis$cluster_comparaison)

input.matrix <- data.matrix(logit98a13_conf_matrix5)

colnames(input.matrix) = c("1", "2", "3", "4", "5", "6", "7", "8")

rownames(input.matrix) = colnames(input.matrix)

confusion5 <- as.data.frame(as.table(input.matrix))

pdf("Matrice de confusion modèle d'affectation 1998 à 2013 5.pdf",
    paper='A4r', width=1500)

plot <- ggplot(confusion5)

plot + geom_tile(aes(x=Var1, y=Var2, fill=Freq)) +
  scale_x_discrete(name="Type de ménage prédit") + scale_y_discrete(name="Type
de ménage observé") + scale_fill_gradient(breaks=seq(from=-.5, to=4, by=.2))
+ labs(fill="Fréquence\nnormalisée") + theme(legend.text =
  element_text(size=18), legend.title = element_text(size=20), axis.title.x =
  element_text(size=20), axis.title.y = element_text(size=20), axis.text.x =
  element_text(size=15), axis.text.y = element_text(size=15))

```

```

dev.off()

logit98a13_conf_matrix6<-table(OD98a13_MTQ_logit_bis$predict6,
OD98a13_MTQ_logit_bis$cluster_comparaison)

input.matrix <- data.matrix(logit98a13_conf_matrix6)

colnames(input.matrix) = c("1", "2", "3", "4", "5", "6", "7", "8")

rownames(input.matrix) = colnames(input.matrix)

confusion6 <- as.data.frame(as.table(input.matrix))

pdf("Matrice de confusion modèle d'affectation 1998 à 2013 6.pdf",
paper='A4r', width=1500)

plot <- ggplot(confusion6)

plot + geom_tile(aes(x=Var1, y=Var2, fill=Freq)) +
scale_x_discrete(name="Type de ménage prédit") + scale_y_discrete(name="Type
de ménage observé") + scale_fill_gradient(breaks=seq(from=-.5, to=4, by=.2))
+ labs(fill="Fréquence\nnormalisée") + theme(legend.text =
element_text(size=18), legend.title = element_text(size=20), axis.title.x =
element_text(size=20), axis.title.y = element_text(size=20), axis.text.x =
element_text(size=15), axis.text.y = element_text(size=15))

dev.off()

logit98a13_conf_matrix7<-table(OD98a13_MTQ_logit_bis$predict7,
OD98a13_MTQ_logit_bis$cluster_comparaison)

input.matrix <- data.matrix(logit98a13_conf_matrix7)

colnames(input.matrix) = c("1", "2", "3", "4", "5", "6", "7", "8")

rownames(input.matrix) = colnames(input.matrix)

confusion7 <- as.data.frame(as.table(input.matrix))

pdf("Matrice de confusion modèle d'affectation 1998 à 2013 7.pdf",
paper='A4r', width=1500)

plot <- ggplot(confusion7)

plot + geom_tile(aes(x=Var1, y=Var2, fill=Freq)) +
scale_x_discrete(name="Type de ménage prédit") + scale_y_discrete(name="Type

```

```
de ménage observé") + scale_fill_gradient(breaks=seq(from=-.5, to=4, by=.2))
+ labs(fill="Fréquence\nnormalisée") + theme(legend.text =
element_text(size=18), legend.title = element_text(size=20), axis.title.x =
element_text(size=20), axis.title.y = element_text(size=20), axis.text.x =
element_text(size=15), axis.text.y = element_text(size=15))

dev.off()
```

```
logit98a13_conf_matrix8<-table(OD98a13_MTQ_logit_bis$predict8,
OD98a13_MTQ_logit_bis$cluster_comparaison)

input.matrix <- data.matrix(logit98a13_conf_matrix8)

colnames(input.matrix) = c("1", "2", "3", "4", "5", "6", "7", "8")

rownames(input.matrix) = colnames(input.matrix)

confusion8 <- as.data.frame(as.table(input.matrix))

pdf("Matrice de confusion modèle d'affectation 1998 à 2013 8.pdf",
paper='A4r', width=1500)

plot <- ggplot(confusion8)

plot + geom_tile(aes(x=Var1, y=Var2, fill=Freq)) +
scale_x_discrete(name="Type de ménage prédit") + scale_y_discrete(name="Type
de ménage observé") + scale_fill_gradient(breaks=seq(from=-.5, to=4, by=.2))
+ labs(fill="Fréquence\nnormalisée") + theme(legend.text =
element_text(size=18), legend.title = element_text(size=20), axis.title.x =
element_text(size=20), axis.title.y = element_text(size=20), axis.text.x =
element_text(size=15), axis.text.y = element_text(size=15))

dev.off()
```

```
logit98a13_conf_matrix9<-table(OD98a13_MTQ_logit_bis$predict9,
OD98a13_MTQ_logit_bis$cluster_comparaison)

input.matrix <- data.matrix(logit98a13_conf_matrix9)

colnames(input.matrix) = c("1", "2", "3", "4", "5", "6", "7", "8")

rownames(input.matrix) = colnames(input.matrix)

confusion9 <- as.data.frame(as.table(input.matrix))
```

```

pdf("Matrice de confusion modèle d'affectation 1998 à 2013 9.pdf",
paper='A4r', width=1500)

plot <- ggplot(confusion9)

plot + geom_tile(aes(x=Var1, y=Var2, fill=Freq)) +
scale_x_discrete(name="Type de ménage prédit") + scale_y_discrete(name="Type
de ménage observé") + scale_fill_gradient(breaks=seq(from=-.5, to=4, by=.2))
+ labs(fill="Fréquence\nnormalisée") + theme(legend.text =
element_text(size=18), legend.title = element_text(size=20), axis.title.x =
element_text(size=20), axis.title.y = element_text(size=20), axis.text.x =
element_text(size=15), axis.text.y = element_text(size=15))

dev.off()

logit98a13_conf_matrix10<-table(OD98a13_MTQ_logit_bis$predict10,
OD98a13_MTQ_logit_bis$cluster_comparaison)

input.matrix <- data.matrix(logit98a13_conf_matrix10)

colnames(input.matrix) = c("1", "2", "3", "4", "5", "6", "7", "8")

rownames(input.matrix) = colnames(input.matrix)

confusion10 <- as.data.frame(as.table(input.matrix))

pdf("Matrice de confusion modèle d'affectation 1998 à 2013 10.pdf",
paper='A4r', width=1500)

plot <- ggplot(confusion10)

plot + geom_tile(aes(x=Var1, y=Var2, fill=Freq)) +
scale_x_discrete(name="Type de ménage prédit") + scale_y_discrete(name="Type
de ménage observé") + scale_fill_gradient(breaks=seq(from=-.5, to=4, by=.2))
+ labs(fill="Fréquence\nnormalisée") + theme(legend.text =
element_text(size=18), legend.title = element_text(size=20), axis.title.x =
element_text(size=20), axis.title.y = element_text(size=20), axis.text.x =
element_text(size=15), axis.text.y = element_text(size=15))

dev.off()

```

ANNEXE F : VALEURS P POUR LE TEST DE KOLMOGOROV-SMIRNOV SUR LE NOMBRE DE DÉPLACEMENTS



ANNEXE G : VALEURS P POUR LE TEST DE KOLMOGOROV-SMIRNOV SUR LA PART DE DÉPLACEMENTS EN POINTE MATINALE



ANNEXE H : VALEURS P POUR LE TEST DE KOLMOGOROV-SMIRNOV SUR LA PART DE DÉPLACEMENTS EN AUTO



ANNEXE I : VALEURS P POUR LE TEST DE KOLMOGOROV-SMIRNOV SUR LES CHÂÎNES COMPLEXES



ANNEXE J : MATRICES DE CONFUSION DES 10 SIMULATIONS DU MODÈLE D'AFFECTATION

Simulation 1	1	2	3	4	5	6	7	8	Total prédicts
1	13946	6487	331	1660	13211	18564	13311	1997	69507
2	2838	40373	15034	6157	20736	4776	9492	2978	102384
3	5	9081	58783	8709	1829	5	277	3633	82322
4	416	10331	5796	1814	4727	679	1878	865	26506
5	16230	18220	1413	3543	25071	27485	20908	3040	115910
6	18094	6081	179	1960	25591	45618	28560	3997	130080
7	15283	7383	329	2104	20501	29264	21676	3038	99578
8	2438	4317	485	729	4332	3659	3414	590	19964
Total observés	69250	102273	82350	26676	115998	130050	99516	20138	646251

Simulation 2	1	2	3	4	5	6	7	8	Total prédicts
1	13990	6308	297	1595	13050	18728	13327	1973	69268
2	2893	40207	15051	6176	20799	4735	9424	2983	102268
3	8	8975	58911	8625	1812	16	268	3647	82262
4	435	10435	5724	1881	4664	677	1800	842	26458
5	15936	18192	1364	3519	25320	27213	21116	3085	115745
6	18111	6148	178	2074	25393	45742	28656	3994	130296
7	15384	7565	321	2069	20534	29346	21491	3069	99779
8	2493	4443	504	737	4426	3593	3434	545	20175
Total observés	69250	102273	82350	26676	115998	130050	99516	20138	646251

Simulation 3	1	2	3	4	5	6	7	8	Total prédicts
1	13998	6424	319	1626	13139	18378	13298	1939	69121
2	2948	40533	15075	6021	20532	4717	9380	2982	102188
3	5	9088	58732	8692	1898	12	267	3617	82311
4	462	10345	5839	1890	4923	703	1861	829	26852
5	16115	18146	1361	3592	25318	27337	20846	3242	115957
6	18009	5936	204	2011	25332	45872	28918	3952	130234
7	15222	7460	319	2090	20486	29305	21544	3063	99489
8	2491	4341	501	754	4370	3726	3402	514	20099
Total observés	69250	102273	82350	26676	115998	130050	99516	20138	646251

Simulation 4	1	2	3	4	5	6	7	8	Total prédicts
1	13908	6378	327	1629	13157	18472	13237	1926	69034
2	2927	40227	15037	6042	20834	4882	9415	3028	102392
3	4	9010	58841	8745	1834	16	292	3568	82310
4	440	10367	5785	1863	4683	684	1793	898	26513
5	16118	18528	1425	3503	25203	27201	21103	3113	116194
6	18183	5963	198	2043	25407	45867	28533	3992	130186
7	15173	7522	296	2114	20390	29329	21663	3093	99580
8	2497	4278	441	737	4490	3599	3480	520	20042
Total observés	69250	102273	82350	26676	115998	130050	99516	20138	646251

Simulation 5	1	2	3	4	5	6	7	8	Total prédicts
1	13932	6582	285	1614	13249	18745	13411	1943	69761
2	3024	39983	15031	6085	20483	4750	9408	3048	101812
3	6	9160	58667	8653	1825	14	259	3590	82174
4	441	10417	5917	1838	4800	657	1888	896	26854
5	16048	18197	1438	3498	25519	27288	21168	3114	116270
6	18230	6046	221	2026	25408	45766	28645	3803	130145
7	15129	7493	329	2186	20231	29160	21368	3188	99084
8	2440	4395	462	776	4483	3670	3369	556	20151
Total observés	69250	102273	82350	26676	115998	130050	99516	20138	646251

Simulation 6	1	2	3	4	5	6	7	8	Total prédicts
1	13757	6347	322	1661	13245	18814	13443	1977	69566
2	2976	40493	15234	6170	20845	4832	9257	3064	102871
3	7	9062	58794	8603	1861	12	252	3632	82223
4	484	10375	5631	1879	4683	671	1876	862	26461
5	16096	18190	1397	3546	25064	27123	21308	3019	115743
6	18186	6093	186	1993	25492	45634	28376	3886	129846
7	15345	7361	342	2105	20388	29257	21734	3169	99701
8	2399	4352	444	719	4420	3707	3270	529	19840
Total observés	69250	102273	82350	26676	115998	130050	99516	20138	646251

Simulation 7	1	2	3	4	5	6	7	8	Total prédits
1	13647	6362	333	1615	13311	18580	13432	1966	69246
2	2867	40514	15144	6182	20671	4737	9513	3057	102685
3	6	9028	58747	8637	1806	13	262	3632	82131
4	428	10478	5764	1892	4755	653	1806	847	26623
5	16423	18073	1367	3472	25081	27436	20993	2998	115843
6	18093	5940	186	2069	25466	45608	28317	3999	129678
7	15380	7411	311	2122	20460	29338	21702	3113	99837
8	2406	4467	498	687	4448	3685	3491	526	20208
Total observés	69250	102273	82350	26676	115998	130050	99516	20138	646251

Simulation 8	1	2	3	4	5	6	7	8	Total prédits
1	13894	6447	281	1579	13152	18544	13374	1865	69136
2	2895	40351	15142	6088	20557	4770	9474	3010	102287
3	6	9066	58702	8757	1868	9	306	3588	82302
4	441	10439	5772	1806	4739	685	1814	891	26587
5	16019	18271	1460	3559	25466	27307	21192	3204	116478
6	18238	6040	206	2046	25159	45668	28479	3925	129761
7	15249	7346	334	2106	20526	29367	21437	3118	99483
8	2508	4313	453	735	4531	3700	3440	537	20217
Total observés	69250	102273	82350	26676	115998	130050	99516	20138	646251

Simulation 9	1	2	3	4	5	6	7	8	Total prédits
1	13859	6482	282	1604	13302	18597	13281	1894	69301
2	2992	39779	14850	6063	20529	4708	9391	3002	101314
3	8	9183	59030	8713	1887	11	322	3593	82747
4	460	10497	5803	1829	4593	662	1851	875	26570
5	16126	18206	1392	3462	25405	27276	21214	3164	116245
6	18133	6248	205	2100	25506	45763	28396	3889	130240
7	15210	7460	343	2186	20298	29419	21624	3167	99707
8	2462	4418	445	719	4478	3614	3437	554	20127
Total observés	69250	102273	82350	26676	115998	130050	99516	20138	646251

Simulation 10	1	2	3	4	5	6	7	8	Total prédicts
1	13765	6397	297	1716	13290	18666	13339	1965	69435
2	2897	40349	15012	6054	20668	4749	9373	3006	102108
3	8	8993	58822	8658	1846	11	274	3620	82232
4	485	10368	5824	1894	4826	618	1923	815	26753
5	16107	18430	1366	3502	25250	27476	21140	3120	116391
6	18298	6004	215	2127	25139	45560	28370	3934	129647
7	15204	7364	348	2033	20534	29238	21658	3129	99508
8	2486	4368	466	692	4445	3732	3439	549	20177
Total observés	69250	102273	82350	26676	115998	130050	99516	20138	646251