

Travail de fin d'études / Projet de fin d'études : La multimodalité dans les transports, réalité ou illusion ? Le cas des Etats-Unis.

Auteur : Bois, Pierre-Antoine

Promoteur(s) : Cools, Mario

Faculté : Faculté des Sciences appliquées

Diplôme : Master : ingénieur civil architecte, à finalité spécialisée en "urban and environmental engineering"

Année académique : 2023-2024

URI/URL : <http://hdl.handle.net/2268.2/19776>

Avertissement à l'attention des usagers :

Tous les documents placés en accès ouvert sur le site le site MatheO sont protégés par le droit d'auteur. Conformément aux principes énoncés par la "Budapest Open Access Initiative"(BOAI, 2002), l'utilisateur du site peut lire, télécharger, copier, transmettre, imprimer, chercher ou faire un lien vers le texte intégral de ces documents, les disséquer pour les indexer, s'en servir de données pour un logiciel, ou s'en servir à toute autre fin légale (ou prévue par la réglementation relative au droit d'auteur). Toute utilisation du document à des fins commerciales est strictement interdite.

Par ailleurs, l'utilisateur s'engage à respecter les droits moraux de l'auteur, principalement le droit à l'intégrité de l'oeuvre et le droit de paternité et ce dans toute utilisation que l'utilisateur entreprend. Ainsi, à titre d'exemple, lorsqu'il reproduira un document par extrait ou dans son intégralité, l'utilisateur citera de manière complète les sources telles que mentionnées ci-dessus. Toute utilisation non explicitement autorisée ci-avant (telle que par exemple, la modification du document ou son résumé) nécessite l'autorisation préalable et expresse des auteurs ou de leurs ayants droit.



Faculté des Sciences Appliquées

TRAVAIL DE FIN D'ÉTUDE

-

*La multimodalité dans les transports,
réalité ou illusion ? Le cas des Etats-Unis*

Auteur : M. Pierre-Antoine BOIS (ingénieur-architecte)

Promoteur : M. Mario COOLS

Membres du jury : M. Mario COOLS, M. Jacques TELLER, Mme. Sabine LIMBOURG

Formation : Master Ingénieur Civil Architecte – Université de Liège

Année académique : 2023-2024

RÉSUMÉ

Le premier objectif de ce travail de fin d'étude consiste, d'une part, à évaluer la part des personnes multimodales dans la population américaine. D'autre part, à mettre en lumière les facteurs influençant la multimodalité des usagers (facteurs socio-économiques, liés aux moyens de transports, liés au contexte ou encore au comportement et à la perception des transports). Les données utilisées sont celles collectées par le département des Transports des États-Unis (USDOT) sur l'année 2022 au terme d'une enquête nationale (*National Household Travel Surveys* ou *NHTS*).

Cette étude, conduite auprès de 10591 personnes, comptabilise 31074 déplacements (utilisation d'un mode de déplacement à la fois) et donne des informations très diverses, tant sur le déplacement (facteurs liés au transport et reconstitution du trajet) que sur les personnes interrogées (facteurs sociaux-économiques). En revanche, les données fournies ne permettent pas de tirer des conclusions sur les questions liées à la perception des transports ou encore le comportement même des personnes (facteurs psychologiques et culturels entre autres). Les données brutes ont été retraitées, restructurées puis analysées notamment via un modèle de régression logistique spécifiquement développé pour ce travail de recherche.

Les résultats montrent que, sur la base de critères arbitraires permettant d'identifier une personne comme multimodale, la population américaine est globalement très peu sujette à la multimodalité. De plus, certains facteurs sociaux économiques comme l'âge, le fait de conduire ou le genre ont une influence importante sur la multimodalité des personnes sondées. La multimodalité se manifeste davantage chez les personnes jeunes qui ne sont pas de sexe féminin et qui conduisent. D'autres facteurs liés au contexte du déplacement ont également une influence sur la multimodalité comme la distance, le temps de trajet, la densité urbaine, le jour de la semaine ou encore le but du voyage. Là encore, la multimodalité est plus présente lors des trajets en semaine, sur une période et une distance relativement courte, au sein des zones densément peuplées. Elle concerne en particulier les personnes se déplaçant pour atteindre un autre mode de transport ou pour accéder à des activités non essentielles comme les loisirs.

ABSTRACT

The first objective of this end-of-study thesis is, on the one hand, to evaluate the share of multimodal people in the American population. On the other hand, to highlight the factors influencing the multimodality of users (socio-economic factors, related to means of transport, linked to the context or to the behavior and perception of transport). The data used are those collected by the United States Department of Transportation (USDOT) for the year 2022 at the end of a national survey (*National Household Travel Surveys or NHTS*).

This study, conducted among 10591 people, counts 31074 trips (use of one mode of travel at a time) and gives a wide range of information, both on the trip (factors related to transport and reconstruction of the journey) and on the people interviewed (social and economic factors). On the other hand, the data provided do not allow conclusions to be drawn on issues related to the perception of transport or the behavior of people themselves (psychological and cultural factors among others). The raw data were reprocessed, restructured and then analyzed, in particular via a logistic regression model specifically developed for this research work.

The results show that, based on arbitrary criteria for identifying a person as "multimodal", the American population is generally very little subject to multimodality. In addition, certain socio-economic factors such as age, driving or gender have a significant influence on the multimodality of the people surveyed. Multimodality is more prevalent among young people who are not female and who drive. Other factors related to the context of travel also have an influence on multimodality, such as distance, travel time, urban density, day of the week or purpose of the trip. Here again, multimodality is more present during weekday journeys, over a relatively short period and distance, in densely populated areas. It concerns in particular people traveling to reach another mode of transport or to access non-essential activities such as leisure.

REMERCIEMENTS

Je tiens ici à remercier chaleureusement monsieur Mario Cools, mon promoteur, qui m'a accompagné tout au long de ce travail de fin d'étude. Je le remercie tout d'abord pour sa disponibilité et sa grande adaptabilité afin de pouvoir organiser des revues très régulièrement et alors que celui-ci était très sollicité. Je tiens également à le remercier pour le suivi global de mon travail et de la pertinence de ses retours critiques et conseils d'une façon générale. M. Cools s'est personnellement beaucoup impliqué dans ce TFE afin que je puisse rapidement monter en compétence dans des domaines très techniques comme les statistiques avancées ou encore la programmation sur le logiciel R par exemple. Dans le même temps, celui-ci m'a accordé beaucoup de liberté et d'autonomie tout au long du TFE, ce qui a été particulièrement agréable pour moi, qui préfère un suivi sur mesure plutôt qu'un encadrement trop rigide. Étant plus attiré par les l'ingénierie que par l'architecture, j'ai dès le départ cherché un sujet lié aux sciences appliquées au bâtiment ou à l'urbanisme et M. Cools m'a tout de suite séduit avec le sujet de ce TFE. Je n'étais au départ ni expert en statistiques ni en programmation mais l'accompagnement dont j'ai bénéficié m'a permis de monter rapidement en compétence et ainsi de prendre véritablement plaisir à explorer et à analyser les données mises à disposition.

Je tenais à également remercier les membres de mon jury : monsieur Jacques Teller et madame Sabine Limbourg, qui ont accepté sans hésitation de suivre mon travail. Enfin, j'adresse un grand merci à ma famille et à mes amis, qui m'ont soutenu sans relâche pendant la durée du TFE, m'apportant conseils et critiques, notamment durant la relecture de mon travail.

TABLE DES MATIÈRES

RÉSUMÉ	1
ABSTRACT	2
REMERCIEMENTS.....	3
1- INTRODUCTION	6
2- ETAT DE L'ART	8
2.1 Facteurs d'influence	8
2.2 Facteurs socio-économiques	10
2.3 Facteurs spécifiques au mode de transport	14
2.4 Facteurs liés au contexte	18
2.5 Facteurs psychologiques.....	23
3- ANALYSE.....	28
3.1 Source de Données.....	28
3.1.1 NHTS 2022, présentation.....	28
3.1.2 Collecte des données.....	30
3.2 Traitement des données.....	33
3.2.1 Mise en forme et filtrage des données de base	33
3.2.2 Prétraitement des données et définition de la multimodalité	36
3.2.3 Ajustement des variables observées	39
3.2.4 Préparation spécifique en vue de la régression logistique.....	41
3.3 Variables retenues et tests statistiques préliminaires.....	42
3.4 Régression logistique et corrélation entre les variables.....	45
3.4.1 Principes et fonctionnement du modèle.....	45
3.4.2 Multicolinéarité et lien entre variables	47
3.4.3 Sélection des variables d'intérêt	48
3.5 Pertinence et limites du modèle	50
3.5.1 ROC.....	51
3.5.2 Hosmer-Lemeshow.....	51
3.5.3 MacFadden	52
3.5.4 Table de classification	53
3.5.5 Limites de la régression logistique	54
3.5.6 Limites du logiciel R.....	55
3.6 Résultats et discussion.....	56
3.6.1 Proportion de multimodaux	56
3.6.2 Facteurs influençant la multimodalité.....	56
4- CONCLUSION	67

5- ANNEXES.....	70
5.1 Compléments	70
5.1.1 Figures	70
5.1.2 Tableaux.....	70
5.1.3 Programmation sur R (code).....	82
5.2 Sources	106

1- INTRODUCTION

La multimodalité est définie par le Larousse en ligne comme « l'utilisation combinée de plusieurs modes de transport au cours d'un même trajet » (Larousse, 2024). Cette définition pose dès le départ plusieurs questions notamment sur ce que l'on entend par la notion de trajet ou encore par la définition d'une combinaison de modes de transports. Dans la littérature, la définition de trajet varie d'un auteur à l'autre car certains le définissent comme la somme des déplacements durant une période donnée (par exemple la journée), d'autres l'utilisent pour parler du déplacement. Le déplacement est aussi ambigu car, selon les sources, il peut faire référence au fait d'aller d'un point A à un point B sans changer de mode de transport, dans d'autres cas, il fait référence à un trajet comme la somme des déplacements sur une période donnée.

Une combinaison de modes de transports implique généralement un ou plusieurs changements de modes de transports lors d'un trajet. Certains modes sont clairement identifiés par la littérature comme étant des modes de transport à part entière (ex : les transports en commun ou les véhicules), d'autres font beaucoup plus débat comme la marche qui n'est pas toujours considérée comme un mode propre. Ce mode fait généralement le lien entre différents modes de transport (par exemple pour passer de la voiture au train) ou constitue un mode de déplacement privilégié (dans le cas d'une balade ou encore pour faire ses courses).

L'utilisation des transports en commun est souvent considérée par la littérature comme intermodale (Ralph Buehler & al., 2007), car les individus accèdent généralement aux arrêts ou aux stations de transport en commun à pied, à vélo ou en voiture. La somme des combinaisons de transport sur une journée constitue une séquence modale, ou chaîne de voyage d'après les précédents auteurs, qui peut être parfois complexe (ex : utilisation du vélo puis des transports en communs pour finir par le covoiturage).

Ainsi, la multimodalité est avérée lorsqu'un individu utilise plus d'un mode de transport au cours d'une période donnée d'après Zihao An & al. (2022). Cela implique de définir certains nombres de critères pour qualifier ou non un trajet comme étant multimodal. Une personne se rendant à la gare en voiture et utilisant deux modes différents (la voiture et le train, sans compter la marche si on la considère) est donc tout aussi multimodale qu'une personne changeant de quais dans une gare afin de passer du tramway au train de banlieues. Le caractère multimodal d'un trajet dépend également de la notion de temps de transport et du temps d'attente ou de transfert entre les différents modes (*access and egress*).

Un temps d'attente trop long entre deux modes (par exemple à un arrêt de bus ou une gare) peut créer une discontinuité à l'échelle du trajet qui peut alors être considéré comme deux trajets distincts avec chacun leur propre séquence de déplacement. Certains auteurs comme Ralph Buehler & al. (2007) montrent que la définition d'un seuil, en fonction du pourcentage de déplacements effectués par un seul mode de transport, est parfois nécessaire pour distinguer des tendances monomodales (un seul mode de transport) de tendances multimodales (plusieurs modes).

La multimodalité constitue un enjeu majeur dans un monde qui cherche à trouver des alternatives crédibles aux véhicules individuels tout en facilitant au maximum les déplacements des personnes. D'après Stephan Krygsman & al. (2004), des mesures politiques et de planification bien élaborées permettraient de minimiser l'influence de certains maillons faibles dans les chaînes de transport public multimodales. La voiture reste un moyen privilégié de déplacement, en particulier dans les sociétés occidentales, et les alternatives doivent être développées pour éviter la friction dans la chaîne modale (ou séquence de déplacements).

D'après ces mêmes auteurs, des efforts ont été récemment effectués par les pouvoirs publics pour une meilleure planification des transports (meilleur tracé des voies d'accès, emplacement des arrêts revu, etc.), sans vraiment prendre en compte les pertes de temps liées transfert d'un mode à l'autre. Un temps de transfert modal trop long réduit considérablement la multimodalité et encourage au contraire l'utilisation des véhicules personnels.

L'étude de la multimodalité chez les usagers passe également par l'analyse de plusieurs facteurs de différentes natures décrivant l'utilisateur même ou encore le mode de transport emprunté. Les facteurs socio-économiques donnent de précieuses informations sur les usagers même, en collectant des données propres à chacun telles que l'âge ou encore le niveau de revenu. Les facteurs liés à la question du mode de transport, impliquent généralement de connaître certains éléments clés comme le temps de transport ou la distance parcourue.

Par ailleurs, d'autres facteurs liés au contexte du voyage (comme le moment de la journée ou encore le but du voyage) sont tout aussi intéressants pour déterminer dans quelle situation la multimodalité se manifeste le plus chez les usagers. Tous ces facteurs précédemment décrits sont en général bien documentés dans la littérature mais peu de recherches prennent en compte les comportements multimodaux ou la perception des transports afin notamment de comprendre la direction et l'ampleur des potentiels déplacements modaux, ce que défend par exemple [Marco Diana & al. \(2009\)](#).

Des outils comme le *MaaS (Mobility as a Service)* sont présentés comme des systèmes intégrés intelligents d'information, de réservation, d'achat et de validation des services de mobilité, centrés sur l'utilisateur permettant d'offrir un service optimal à chaque utilisateur ([Smart City Institute, 2021](#)). Des études approfondies doivent être conduites pour mieux identifier les comportements des usagers afin de pouvoir proposer plus de combinaisons multimodales et ainsi lutter contre le phénomène de monomodalité, souvent associé à une utilisation abusive de la voiture individuelle. Reste que, malgré un développement important d'alternatives crédibles au transport privé ces dernières années, certains auteurs, comme [Eva Heinen & Kiron Chatterjee \(2015\)](#), constatent que la voiture est très largement utilisée dans le cas de déplacements monomodaux et est même un mode de transport très utilisé dans beaucoup de séquences multimodales (ex : pour se rendre à la gare).

Face à ce constat alarmant, l'objectif de ce travail est dans un premier d'évaluer la proportion de personnes multimodales au sein de la population étudiée, à savoir la population américaine. Dans un second, il s'agit de déterminer les facteurs influençant le plus la multimodalité des usagers, qu'ils soient de nature socio-économiques, liées aux transports ou au contexte, ou encore d'ordre comportemental. Ce travail a pour but de répondre à la question suivante : La multimodalité aux Etats-Unis, réalité ou illusion ?.

Pour ce faire, nous commencerons par un état de l'art présentant les différents facteurs qui ont un impact sur la multimodalité en les classant par type de facteur. Dans un second temps, nous présenterons la source des données utilisées pour répondre à ces questions ainsi que la méthodologie de l'étude. Enfin, nous analyserons ces données, et les résultats seront alors discutés en lien avec la littérature.

2- ETAT DE L'ART

2.1 Facteurs d'influence

Plusieurs types de facteurs ont une influence sur la multimodalité des personnes et ils peuvent être réunis en quatre grandes catégories : les facteurs socio-économiques, les facteurs liés au modes de transport, ceux liés au contexte et enfin ceux liés à la perception des transports et au comportement des usagers. Les trois premières catégories incorporent des facteurs dits *objectifs* qui sont facilement observables et quantifiables, à l'inverse des facteurs liés au comportement et à la perception des usagers qui sont considérés comme *subjectifs*.

Tableau 1 : Facteurs influençant la multimodalité classés par type

Type de facteur d'influence	Études qui en témoignent
Facteurs socio-économiques	
Genre	Ralph Buehler & al., 2007 ; Stephan Krygsman & al., 2004 ; Eva Heinen & Kiron Chatterjee, 2015 ; Eric J.E. Molin & al., 2010 ; Thomas Klinger, 2017 ; Heinen & Mattioli, 2019 ; Sungyop Kim & al., 2007 ; Blumenberg & Pierce, 2013 ; Zhang et al., 2014
Permis de conduire	Stephan Krygsman & al., 2004 ; Eva Heinen & Kiron Chatterjee, 2015 ; Heinen & Mattioli, 2019 ; Bai & Kattan, 2014
Possession d'une voiture	Ralph Buehler & al., 2007 ; Eva Heinen & Kiron Chatterjee, 2015 ; Zihao An & al., 2021 ; Marco Diana, 2012 ; Bai & Kattan, 2014 ; Blumenberg & Pierce, 2013
Possession d'un vélo	Stephan Krygsman & al., 2004 ; Eric J.E. Molin & al., 2010
Âge	Ralph Buehler & al., 2007 ; Stephan Krygsman & al., 2004 ; Eva Heinen & Kiron Chatterjee, 2015 ; Heinen & Mattioli, 2019 ; Zihao An & al., 2021 ; Yang Jiang & al., 2012 ; Sungyop Kim & al., 2007 ; Blumenberg & Pierce, 2013 ; Zhang et al., 2014
Situation familiale	Ralph Buehler & al., 2007 ; Thomas Klinger, 2017
Revenu	Eva Heinen & Kiron Chatterjee, 2015 ; Thomas Klinger, 2017 ; Hye Kyung Lee & al., 2021 ; Heinen & Mattioli, 2019 ; Chieh-Hua Wen & al., 2012 / Yang Jiang & al., 2012 ; Sungyop Kim & al., 2007 ; Blumenberg & Pierce, 2013 ; Zhang et al., 2014
Taille du logement	Eva Heinen & Kiron Chatterjee, 2015
Statut salarial	Ralph Buehler & al., 2007 ; Eva Heinen & Kiron Chatterjee, 2015 ; Eric J.E. Molin & al., 2010 ; Thomas Klinger, 2017 ; Zihao An & al., 2021 ; Blumenberg & Pierce, 2013
Niveau d'éducation	Ralph Buehler & al., 2007 ; Eva Heinen & Kiron Chatterjee, 2015 ; Marco Diana, 2012 ; Sungyop Kim & al., 2007 ; Blumenberg & Pierce, 2013 ; Zhang et al., 2014
Origines ethniques	Ralph Buehler & al., 2007 ; Heinen & Mattioli, 2019 ; Blumenberg & Pierce, 2013
Facteurs liés aux transports	
Distance	Stephan Krygsman & al. ; 2004, Eric J.E. Molin & al., 2010 ; Laura Alessandretti & al., 2022 ; Sungyop Kim & al., 2007 ; Blumenberg & Pierce, 2013 ; Zhang et al., 2014
Temps de trajet	Stephan Krygsman & al., 2004 ; Riccardo Gallotti & Marc Barthelemy, 2014 ; Ehab I. Diab & Ahmed M. El-Geneidy, 2012

Transfert/correspondance	Stephan Krygsman & al., 2004 ; Li et al., 2024 ; Riccardo Gallotti & Marc Barthelemy, 2014 ; Ehab I. Diab & Ahmed M. El-Geneidy, 2012 ; Polydoropoulou & Ben-AkivaA., 2001 ; Duthilleul et al., 1999 ; Yuji Sh, 2023 ; Zhang et al., 2014
Accès facilité aux transports	Stephan Krygsman & al., 2004 ; Eric J.E. Molin & al., 2010 ; Haochun & Yunyi, 2023 ; Laura Alessandretti & al., 2022 ; Yang Jiang & al., 2012 ; Sungyop Kim & al., 2007 ; Zhang et al., 2014
Transition entre les modes	Eric J.E. Molin & al., 2010 ; Polydoropoulou & Ben-AkivaA., 2001 ; Blumenberg & Pierce, 2013
Intéactions entre les modes (interconnectivité)	Ralph Buehler & al., 2007 ; Eva Heinen & Kiron Chatterjee, 2015 ; Chieh-Hua Wen & al., 2012 ; Yang Jiang & al., 2012 ; Zhang et al., 2014
Réseaux de transport	Xavier Bach & al., 2023 ; Laura Alessandretti & al., 2022
Nouvelles alternatives	Aoyong Li & al., 2022 ; Felix Schwinger & al., 2022
Approche nodale	Yifan Yue & al., 2021 ; Xavier Bach & al., 2023
Information en temps réelle	Nguyen Hoang-Tung & al., 2021 ; Chorus et al., 2007 ; Bai & Kattan, 2014
Influence du contexte	
Moment de la journée	Stephan Krygsman & al., 2004 ; Eric J.E. Molin & al., 2010 / Bai & Kattan, 2014
Fréquence (quotidienne et hebdomadaire)	Ralph Buehler & al., 2007 ; Eva Heinen & Kiron Chatterjee, 2015 ; Bai & Kattan, 2014 ; Blumenberg & Pierce, 2013
Retards	Chieh-Hua Wen & al., 2012 ; Duthilleul et al., 1999 ; Yuji Sh, 2023 ; Ehab I. Diab & Ahmed M. El-Geneidy, 2012 ; Laura Alessandretti & al., 2022 ; Bai & Kattan, 2014
Coût	Marco Diana & al., 2009 ; Eric J.E. Molin & al., 2010 ; Chieh-Hua Wen & al., 2012
Commodité du système	Stephan Krygsman & al., 2004 ; Eric J.E. Molin & al., 2010 ; Polydoropoulou & Ben-AkivaA., 2001 ; Ehab I. Diab & al., 2015 ; Chieh-Hua Wen & al., 2012 ; Sungyop Kim & al., 2007 ; Zhang et al., 2014
Qualité/fiabilité des transports	Eric J.E. Molin & al., 2010 ; Duthilleul et al., 1999 ; Ehab I. Diab & al., 2015 ; Yang Jiang & al., 2012 ; Sungyop Kim & al., 2007 ; Zhang et al., 2014
Offre des transports (disponibilité)	Polydoropoulou & Ben-AkivaA., 2001 ; Yuji Sh, 2023
Confort	Eric J.E. Molin & al., 2010 ; Polydoropoulou & Ben-AkivaA., 2001 ; Yang Jiang & al., 2012 ; Sungyop Kim & al., 2007
Criminalité	Ehab I. Diab & al., 2015 ; Sungyop Kim & al., 2007 ; Eric J.E. Molin & al., 2010)
Danger du trafic	Yang Jiang & al., 2012 ; Sungyop Kim & al., 2007 ; Zhang et al., 2014 a
Motif du déplacement	Stephan Krygsman & al., 2004 ; Eric J.E. Molin & al., 2010 ; Chieh-Hua Wen & al., 2012 ; Blumenberg & Pierce, 2013 ; Zhang et al., 2014
Utilisation du sol	Stephan Krygsman & al., 2004 ; Sungyop Kim & al., 2007 ; Blumenberg & Pierce, 2013 ; Zhang et al., 2014
Densité urbaine	Ralph Buehler & al., 2007 ; Stephan Krygsman & al., 2004 ; Yuji Sh, 2023 ; Marco Diana, 2012 ; Sungyop Kim & al., 2007 ; Blumenberg & Pierce, 2013
Météo	Eric J.E. Molin & al., 2010 ; Bai & Kattan, 2014 ; Zhang et al., 2014
Bagages et compagnons de voyages (animaux)	Ralph Buehler & al., 2007 ; Eric J.E. Molin & al., 2010
Voyager avec d'autres personnes	Eric J.E. Molin & al., 2010

Pouvoirs public	Ralph Buehler & al., 2007
Facteurs psychologiques	
Perception du service	Eva Heinen & Kiron Chatterjee, 2015 ; Thomas Klinger, 2017 ; Clauss & Döppe, 2016 ; William jen & Kai-chieh hu, 2003 ; Chorus et al., 2007 ; Zihao An & al., 2021 ; Marco Diana, 2010 ; Marco Diana, 2012 ; Bai & Kattan, 2014
Satisfaction	Zihao An & al., 2022 ; Yongsung Lee et al., 2020 ; Duthilleul et al., 1999 ; Ehab I. Diab & al., 2015 ; Marco Diana, 2010
Culture	Eva Heinen & Kiron Chatterjee, 2015 ; Thomas Klinger, 2017 ; Heinen & Mattioli, 2019 ; Yang Jiang & al., 2012 ; Zhang et al., 2014
Modes de vie et évolution	Ralph Buehler & al., 2007 ; Thomas Klinger, 2017 ; Yongsung Lee et al., 2020 ; Heinen & Mattioli, 2019 ; Marco Diana, 2010 ; Blumenberg & Pierce, 2013
Attitudes et Changements de comportements	Eric J.E. Molin & al., 2010 ; Thomas Klinger, 2017 ; Wokje Abrahamse & al., 2009 ; Yongsung Lee et al., 2020 ; Zihao An & al., 2021 ; Marco Diana, 2010 ; Marco Diana, 2012
Traits de personnalité	Marco Diana, 2010 ; Blumenberg & Pierce, 2013
Dissonance cognitive	Clauss & Döppe, 2016 ; Yang Jiang & al., 2012 ; Bai & Kattan, 2014

2.2 Facteurs socio-économiques

Les facteurs socio-économiques sont des facteurs objectifs donnant des indications sur les individus permettant de les identifier, indépendamment du déplacement. Certains de ces facteurs permettent de mettre en évidence un profil de personne associé à la multimodalité, et de nombreux auteurs ont tenté de les identifier par le passé.

Un premier facteur très important pour différencier les personnes est le genre imputé aux personnes interrogées. Des études conduites par [Stephan Krygsman & al. \(2004\)](#) ou encore [Eric J.E. Molin & al. \(2010\)](#), montrent que les hommes ont tendance à voyager plus longtemps que les femmes en moyenne et sur de plus grandes distances. En revanche, d'après [Ralph Buehler & al. \(2007\)](#), les femmes sont plus multimodales que les hommes, notamment lorsqu'elles prennent les transports publics et semble être imputé par des schémas d'activité plus complexes que chez les hommes, qui tendent à utiliser la voiture plus fréquemment mais de manière monomodale ([Thomas Klinger, 2017](#), [Eva Heinen & Kiron Chatterjee, 2015](#)). Les femmes ont encore aujourd'hui la charge de la plupart des responsabilités ménagères, ce qui implique de voyager davantage que les hommes et donc d'utiliser plus de modes de transport d'après ces derniers auteurs.

Avec l'augmentation de la proportion d'automobilistes multimodaux d'année en année chez les, l'écart entre les deux sexes se réduit progressivement en ce qui concerne les déplacements monomodaux ([Ralph Buehler & al., 2007](#), [Heinen & Mattioli, 2019](#)). Cela est particulièrement vrai aux Etats-Unis où la voiture est omniprésente dans les ménages, ce qui conduit à une utilisation équivalente chez les deux sexes. En Europe au contraire, les femmes sont contraintes de faire plus de déplacements sans voiture (ex : pour aller travailler) car l'homme est en général prioritaire pour utiliser la voiture ([Ralph Buehler & al., 2007](#)). Les hommes sont par ailleurs plus susceptibles de combiner la conduite avec le vélo que les femmes, d'après [Sungyop Kim & al. \(2007\)](#), ce qui peut notamment s'expliquer par un sentiment d'insécurité chez les femme lorsqu'elles utilisent ce mode de transport, notamment la nuit. Ce dernier constat est à nuancer car dans certaines villes en Asie du Sud-Est (ex : Shanghai en Chine), les femmes utilisent autant le vélo que les hommes d'après [Zhang et al. \(2014\)](#), sans doute car

l'insécurité y est moindre que dans d'autres régions du monde. La littérature note également que les hommes sont moins enclins au voyage à des fins récréatives alors qu'au contraire, les femmes tendent à plus utiliser la voiture lors de ce type de voyage, en particulier lorsqu'elles ont de la compagnie (Eric J.E. Molin & al., 2010).

Le permis de conduire est fortement lié à la possession d'une voiture, car un jeune conducteur pourra conduire la voiture de ses parents par exemple sans en avoir la possession. Dans tous les cas, l'utilisation de la voiture est généralement associé à un comportement monomodal qui tend à réduire multimodalité d'après Eva Heinen & Kiron Chatterjee (2015), même si cette tendance tend à se réduire avec le temps d'après certains auteurs comme Heinen & Mattioli (2019). En effet, la voiture est également utilisée pour atteindre un autre mode de transport (ex : une gare, un arrêt de bus, etc.) et elle contribue donc à la multimodalité dans certains cas. D'après Stephan Krygsman & al. (2004), les usagers des transports en commun, qui ont un permis de conduire sont même plus susceptibles de se rendre en station que les usagers qui n'ont pas de permis. La combinaison entre les voitures et les autres moyens de transport est particulièrement intéressante car, malgré un temps de transfert plus important pour changer de mode de transport, cela permet de gagner du temps sur la globalité du trajet (Bai & Kattan, 2014).

Tout comme le permis de conduire, la possession d'une voiture tend à avoir un impact négatif sur la multimodalité des individus (Blumenberg & Pierce, 2013, Eva Heinen & Kiron Chatterjee, 2015) et cette tendance s'accroît même avec l'augmentation du nombre de véhicules dans le foyer. Un accès illimité à un véhicule domestique augmentera en effet considérablement la probabilité que tous les déplacements soient effectués à l'aide d'un véhicule d'après Blumenberg & Pierce (2013). De même, plus une personne a accès à un ou plusieurs véhicules et moins elle sera susceptible d'effectuer des trajets multimodaux, en particulier pour les grandes distances (Ralph Buehler & al., 2007). Selon Sungyop Kim & al. (2007), la dépendance à l'automobile est un phénomène mondial qui tend à augmenter ces dernières années et les gens utilisent une voiture personnelle en raison de sa commodité et de sa flexibilité, même pour de très courts trajets.

Ce phénomène est particulièrement présent en dehors des villes d'après Marco Diana (2012), surtout dans les banlieues où il constitue un mode d'accès privilégié aux gares ferroviaires (Sungyop Kim & al., 2007). On peut cependant noter que, selon Bai & Kattan (2014), le fait d'être propriétaires d'une voiture n'implique pas forcément d'avoir tout le temps accès à son véhicule pour un usage quotidien (ex : voiture empruntée par les enfants) et dans ce cas, l'utilisation de transports en commun augmente. A l'inverse, les automobilistes multimodaux ont tendance à vivre dans des ménages sans voiture selon Ralph Buehler & al. (2007) et sont généralement localisés dans les pays développés (Zihao An & al., 2021). L'accès à des alternatives à la voiture, comme le fait de posséder un vélo, encourage fortement la variabilité modale car les propriétaires de vélos ont la volonté et/ou la nécessité d'utiliser d'autres modes (Eva Heinen & Kiron Chatterjee, 2015).

L'âge est également un facteur déterminant car selon la tranche d'âge considérée, les gens sont plus ou moins multimodaux. Il est communément admis qu'en moyenne, les voyageurs âgés sont moins susceptibles d'être unimodaux (Blumenberg & Pierce, 2013) notamment à cause de la fatigue physique engendrée par le passage d'un mode de transport à l'autre (temps d'attente, marche pour se rendre d'un mode à l'autre, présence d'escaliers, etc.), ce que soutiennent Sungyop Kim & al. (2007). Les voyageurs ont tendance à être en effet moins multimodaux à mesure qu'ils vieillissent (de moins en moins multimodaux après 30 ans), ce qui peut être également imputé à moins de contraintes de travail (retraite) en plus de limitations physiques qui ne favorisent pas la marche. D'autres études indiquent au contraire que les plus de 60 ans semblent marcher aussi loin que les personnes entre 20 à 40 ans. En effet, la notion de temps est différente chez eux (le temps semble s'accélérer) et ces derniers ont

accès à certains privilèges (sièges réservés, billets gratuits, etc.) qui les incitent à utiliser les transports en commun (Yang Jiang & al., 2012). Les seniors retraités sont plus multimodaux que certains les adultes en plein milieu de carrière professionnelle d'après Thomas Klinger (2017) (même s'ils n'atteignent pas le niveau des jeunes adultes) car ils sont détachés des toutes contraintes professionnelles ou familiales.

Les difficultés physiques à utiliser certains modes de transport (tels que les vélos ou la marche) ne constituent pas le seul facteur expliquant cette tendance selon Eva Heinen & Kiron Chatterjee (2015). Cela est également dû à un contexte socio-économique et culturel dans la jeunesse des baby-boomers (nées entre 1946 et 1964), favorisant davantage l'utilisation de la voiture, et ce bien plus tôt en moyenne que les moins de 30 ans. Cela explique donc en partie pourquoi la multimodalité est au plus bas chez les personnes âgées d'après Zihao An & al. (2021). Cependant, les recherches de Eric J.E. Molin & al. (2010) montrent que les voyageurs plus âgés sont beaucoup plus enclins à voyager à des fins récréatives que les jeunes voyageurs notamment lorsqu'ils connaissent bien l'itinéraire vers leur destination finale. En revanche, le temps de trajet est généralement plus long chez les jeunes contrairement aux personnes âgées, sans doute pour des raisons de fatigue physique comme on l'a évoqué (Stephan Krygsman & al., 2004).

Les auteurs constatent une évolution positive de la multimodalité chez les jeunes générations et de façon progressive. Les millenials (nés entre le début des années 1980' et la fin des années 1990', ou >35 ans) sont ainsi, d'après Zihao An & al. (2021) ou encore Thomas Klinger (2017), plus multimodaux que ceux de la génération X (nés entre 1965 et 1979, ou >45 ans) et consacrent beaucoup moins de temps aux voyages en voiture et aux activités de plein air. Ralph Buehler & al. (2007) indiquent que les automobilistes multimodaux ont tendance à être plus jeunes, en particulier les jeunes adultes entre 25 et 32 ans qui, par leur travail, sont plus susceptibles d'être multimodaux que les 16 à 24 ans. D'autres études montrent cependant que même s'il y a globalement une convergence entre les catégories d'âge au cours du temps, même au regard de la multimodalité, les jeunes de moins de 30 ans deviennent de moins en moins multimodaux par rapport aux personnes âgées de 30 à 64 ans (Heinen & Mattioli, 2019). Enfin, on notera que les jeunes adultes (entre 18 à 39 ans) représentent environ 80 % de tous les utilisateurs de vélos électriques dans les grandes villes asiatiques d'après Zhang et al. (2014).

La multimodalité est également liée à la situation familiale, car selon Ralph Buehler & al. (2007) ou encore Thomas Klinger (2017), les personnes sans enfants étaient plus susceptibles d'utiliser leur automobile de façon multimodale, que celles vivant dans un ménage comptant des enfants ou bien plus de deux adultes. D'après ces mêmes auteurs, dans ce dernier cas, les gens ont plus tendance à aller chercher ou à déposer d'autres membres du même ménage, ce qui implique généralement d'utiliser une voiture de façon monomodale pour la plupart des trajets. Ainsi, d'après Stephan Krygsman & al. (2004), les personnes ayant de jeunes enfants choisiront sans doute des lieux de transfert plus proches pour contrer le fardeau supplémentaire de devoir accompagner de jeunes enfants (souvent peu disciplinés et/ou rapidement fatigués). Ce dernier auteur indique également que le temps d'accès sortie aux transports est plus faible chez les célibataires que chez les personnes en couple et leur temps de trajet est généralement plus important.

Le niveau de revenu joue un rôle important sur la multimodalité car d'après Blumenberg & Pierce (2013) ou encore Eva Heinen & Kiron Chatterjee (2015) plus le revenu est élevé et plus la multimodalité des ménages diminue en général. En effet, les personnes à fort revenus privilégient davantage un véhicule personnel car ils ont les moyens de se le payer, ce qui favorise les comportements monomodaux. D'après certains auteurs comme Blumenberg & Pierce (2013), Les déplacements multimodaux sont moins fréquents chez les adultes à faible revenu (moins que 20 000 dollars/an) par rapport aux classes moyennes mais cela reste quand même bien plus de voyages que les voyageurs

unimodaux à revenu élevé. En effet, les voyageurs à faible revenu et non professionnels sont très sensibles aux coûts d'accès aux transports, ce qui peut expliquer en partie pourquoi ils sont moins multimodaux (Chieh-Hua Wen & al., 2012). Reste que selon Blumenberg & Pierce, 2013, Yang Jiang & al. (2012), les personnes à faible revenu utilisent bien plus les transports en commun et la marche que les personnes à revenu élevé. De façon étonnante, les voyageurs des groupes à revenu moyen sont moins susceptibles de se livrer à des voyages multimodaux que les personnes à revenu élevé (supérieur à 100 000 dollars), notamment pour des raisons financières (Blumenberg & Pierce, 2013, Thomas Klinger, 2017).

Heinen & Mattioli (2019) en est venu à la même conclusion en indiquant que les personnes aux revenus intermédiaires et bas ont réduit leur niveau de multimodalité depuis les années 2000 par rapport aux personnes aux revenus élevés, sans doute à cause des récentes crises économiques (ex : crise de 2008). Un accès plus facile à la voiture individuelle, chez les personnes aisées, peut expliquer une tendance plus importante à la multimodalité chez ces derniers, qui n'ont alors pas besoin de marcher pour atteindre un autre moyen de transport (Eva Heinen & Kiron Chatterjee, 2015, Yang Jiang & al., 2012). Dans certaines régions du monde comme en Asie, le vélo est associé à un revenu moyen ou faible (Zhang et al., 2014) alors que dans les grandes villes européennes (Paris, Londres, Amsterdam), il est au contraire vu comme un signe de richesse (proximité avec les activités économiques). Enfin, il faut noter que les personnes à faible revenu vivent en général dans des déserts multimodaux, ou zones de transit qui sont mal connectées aux réseaux de transports, ce qui impacte négativement la multimodalité de ces populations (Hye Kyung Lee & al., 2021).

La notion de revenus est directement corrélée à la taille du logement et il a été démontré par Eva Heinen & Kiron Chatterjee (2015) que plus le logement est petit et plus la multimodalité est faible. Aussi, le fait de vivre dans un appartement par rapport à un logement isolé et d'appartenir aux groupes socio-économiques les plus élevés est associé à un plus grand nombre de modes de transport utilisés.

Dans la continuité de ce qui a été dit concernant les revenus, certains auteurs comme Zihao An & al. (2021) indiquent que le passage d'étudiant à salarié à temps plein augmente le nombre de trajets potentiels (surtout domicile-travail) mais réduit dans le même le nombre de trajets multimodaux. Les travailleurs privilégient en effet les itinéraires les plus courts possibles avec peu de variabilité modales pour éviter tout retard (trajets optimisés et répétitifs). Les auteurs Blumenberg & Pierce, 2013 ou encore Eva Heinen & Kiron Chatterjee (2015) sont arrivés à la même conclusion en ajoutant que les personnes qui travaillent ont tendance à compter sur un seul mode de transport (généralement l'automobile).

Au contraire, les personnes en plein emploi sont moins multimodales que les travailleurs à temps partiel ou les personnes au chômage (Thomas Klinger, 2017, Eva Heinen & Kiron Chatterjee, 2015) et leur temps d'accès de sortie aux transports est plus faible statistiquement (Sungyop Kim & al., 2007). Les personnes sans emploi sont plus susceptibles d'être des automobilistes multimodaux, tant au niveau journalier que hebdomadaire (Ralph Buehler & al., 2007) et ils utilisent alors qu'un seul mode primaire. Cette notion de mode primaire pour accéder au travail est moins visible chez les étudiants qui utilisent les mêmes modes pour différentes activités d'après Eva Heinen & Kiron Chatterjee (2015), qu'il s'agisse de se rendre à l'école ou aux loisirs par exemple.

La multimodalité des personnes est aussi l'affaire d'éducation qui influence grandement le comportement des voyageurs. D'après Marco Diana (2012), les voyageurs instruits ont tendance à vivre plutôt dans les centres-villes là où l'offre multimodale est la plus importante. De même, les automobilistes multimodaux ont tendance à avoir fait des études supérieures à l'école secondaire (Ralph Buehler & al., 2007, Eva Heinen & Kiron Chatterjee, 2015). Mais dans le même temps, selon

Blumenberg & Pierce, (2013), les voyageurs unimodaux sont plus susceptibles d'être instruits ce qui est généralement dû à un accès à de meilleurs emplois et de meilleurs revenus permettant l'achat d'un véhicule personnel. Au contraire, d'après ces mêmes auteurs, les personnes qui n'ont pas fait d'études secondaires, même en tenant compte du revenu, sont moins susceptibles d'utiliser plusieurs modes.

Il faut cependant nuancer en indiquant que des niveaux plus élevés de marche et de vélo ont été constatés chez les personnes ayant un niveau de scolarité plus élevé, en particulier durant les études universitaires où les individus apprennent à apprécier la marche, le vélo et les transports en commun comme des modes de transport viables (Ralph Buehler & al., 2007). Les utilisateurs de bicyclettes sont les personnes ayant un niveau d'éducation moyen ou faible en Asie (Zhang et al., 2014) alors que c'est généralement l'inverse en Occident où le déplacement à vélo est un luxe. Enfin, d'après Eric J.E. Molin & al. (2010), les plus personnes instruites sont un peu plus enclines à voyager avec des bagages lourds et avec d'autres personnes, ou par temps de pluie car elles ont plus tendance à choisir d'autres alternatives que le train.

Concernant l'origine ethnique, Les voyageurs unimodaux sont plus susceptibles d'être blancs non hispaniques selon Blumenberg & Pierce (2013) et que cette tendance augmente au cours du temps d'après Heinen & Mattioli (2019). Toujours d'après ces auteurs, les Afro-Américains ou encore les asiatiques voyagent moins, sont les moins susceptibles d'avoir une automobile et sont plus à même d'effectuer tous leurs déplacements en utilisant un seul mode de transport. Sungyop Kim & al. (2007) ajoutent d'ailleurs que ces populations sont plus susceptibles d'utiliser le bus pour se rendre dans les gares et en revenir car en moyenne leur revenu est généralement plus faible et l'offre de bus est généralement meilleure dans les quartiers à plus faibles revenus (plus d'usagers potentiels). Ces mêmes auteurs précisent que les personnes issues de minorités utilisent d'ailleurs moins souvent la marche à pied dans leur voyage multimodal, sans doute car ils n'habitent pas au cœur des villes dans les zones les plus piétonnisées. Il faut cependant noter que récemment, la proportion d'automobilistes multimodaux a augmenté dans des proportions similaires chez les blancs et comme chez les non-blancs (Ralph Buehler & al., 2007), même si l'écart reste encore très important.

2.3 Facteurs spécifiques au mode de transport

Nous avons identifié des facteurs propres aux usagers qui influencent la multimodalité. Ils existent également d'autres facteurs objectifs associés aux transports qui ont un impact sur la multimodalité.

La majorité des déplacements urbains sont courts et peuvent donc être effectués à pied et viennent ensuite les distances moyennes effectuées, de façon optimale, à vélo ou en bus (Laura Alessandretti & al., 2022). Enfin, les longues distances concernent les déplacements en métro, train (ou autres transports ferroviaires) ou encore en voiture. D'une façon générale, il est admis que les personnes qui parcourent le plus de kilomètres sont plus susceptibles de le faire en utilisant un seul mode de transport, généralement l'automobile (Blumenberg & Pierce, 2013) ou encore les transports en commun (Eric J.E. Molin & al., 2010). L'augmentation de la distance jusqu'à un lieu de transfert réduit considérablement la propension à utiliser les modes lents comme la marche ou le vélo selon Stephan Krygsman & al. (2004) ou encore Eric J.E. Molin & al. (2010), ce qui implique un certain effort physique.

En revanche, selon ces auteurs, la probabilité d'utiliser un vélo a tendance à augmenter par rapport à la marche, dans le cas d'une grande distance effectuée en modes lents, car ce dernier mode fatigue plus. Reste que, comme le soulignent Zhang et al. (2014), le vélo (même électrique) est principalement utilisé pour les trajets sur de courtes et moyennes distances (moins de 10 km). On notera également

que les gens ont tendance à marcher plus loin pour se rendre à une station de train qu'à un arrêt d'autobus (Sungyop Kim & al., 2007).

Indéniablement associé à la distance, le temps lié au trajet a un impact sur la multimodalité. Selon Riccardo Gallotti & Marc Barthelemy (2014), les trajets plus longs profitent de la multimodalité du système et des itinéraires alternatifs peuvent être envisagés lorsque cela est possible en cas de problème sur le trajet. Certains modes seront privilégiés, comme par exemple le taxi, dans le cas de temps de transport importants (Stephan Krygsman & al., 2004), en particulier en sortie de gare ou à un terminus de bus. D'après circulation Ehab I. Diab & Ahmed M. El-Geneydy (2012), la durée du trajet est elle-même influencée par de nombreux facteurs comme distance, le retard au début du voyage, la période de la journée, le nombre d'arrêts réels, les conditions météorologiques entre autres, ce qui a un impact sur les choix des usagers. On comprendra qu'un retard sur une ligne de train poussera davantage un usager de ce type de transport à privilégier d'autres modes comme le taxi ou bien un véhicule personnel.

Dans la même veine, le travail de Riccardo Gallotti & Marc Barthelemy (2014) au Royaume-Uni, indique que près du quart du temps de trajet est perdu dans les correspondances pour les trajets avec plus d'un mode notamment à cause d'un manque de synchronisation entre les modes. D'après les mêmes auteurs, les utilisateurs ont perçu une diminution de leur temps de déplacement alors qu'une réduction du temps de fonctionnement augmente la multimodalité car les usagers sont plus enclins à prendre les transports (Ehab I. Diab & Ahmed M. El-Geneydy, 2012). Toujours selon ces derniers auteurs, des voies réservées aux autobus, l'introduction d'autobus articulés ou encore la mise en place d'autobus à arrêts limités, offrent un service plus rapide qui favorise de fait la multimodalité. Concernant le vélo, la durée moyenne d'un trajet à vélo était de 20 minutes d'après Zhang et al. (2014) et on l'a vu, ce mode est limité aux faibles distances mais peut permettre un gain de temps considérable dans les cœurs de ville très congestionnés.

Le temps de transferts (ou de correspondance) est indissociable du temps de trajet car les usagers utilisant plusieurs types de transport regarderont toujours le temps de trajet total et non les temps de trajet et d'attente pris individuellement. Les auteurs Polydoropoulou & Ben-AkivaA. (2001) alertent sur la nécessité de limiter le temps de transfert dans les systèmes de transport public existants notamment en implémentant des solutions technologiques (ex : panneaux d'information). Pour eux, ce temps de transfert est l'un des freins majeurs au développement de la multimodalité, bien plus que le temps de transport en lui-même. Les correspondances entraînent de facto des temps de transport de ligne beaucoup plus longs (Stephan Krygsman & al., 2004).

Cependant, c'est bien la friction entre les modes secondaires et le mode principal de transport qui poussent la plupart des gens à utiliser un véhicules particulier pour accéder à leur mode principal (ex : la voiture pour se rendre à la gare) plutôt que d'utiliser une combinaison de transports en communs par exemple (Duthilleul et al., 1999). Ce phénomène est d'autant plus accentué par le fait que les gens sont particulièrement affectés par un retard, même léger, par peur de manquer leur correspondance (Li et al., 2024). Les gens préféreront donc utiliser un véhicule personnel que des transports parfois aux horaires variables, voire très en retard. Cependant d'après Yuji Sh (2023), les métros et transports en cœur de ville sont généralement très ponctuels et il est alors bien plus efficace de prendre les transports en commun que d'être congestionné dans le trafic dense des villes.

Concernant l'accès au transports, celui-ci n'est pas le même selon les modes de départ et d'arrivée dans la chaîne multimodale. Le temps d'accès aux transports est globalement bien plus grand pour la marche par exemple que pour le vélo d'après Stephan Krygsman & al. (2004), car le vélo est tout simplement plus rapide et moins fatigant en moyenne que la marche sur une distance donnée. En revanche,

d'après ces mêmes auteurs, le temps d'accès à une gare est considéré comme le même entre les deux modes car il faut considérer le temps pour garer et de verrouiller le vélo et dans ce cas, les deux modes sont aussi compétitifs. Il faut également souligner que la friction est plus importante pour le vélo si on décide de le garder dans les transports, car il faut généralement trouver un espace pour le ranger et car toutes les circulations dans les gares sont dimensionnées pour le cheminement piéton.

Un meilleur aménagement du terrain à proximité des stations de bus ou de trains est essentiel à l'expansion du réseau de vélo d'après [Laura Alessandretti & al. \(2022\)](#) au même titre que l'entretien de ces derniers ou le fait de les mettre en libre-service. De même, un système de tarification plus efficace, le vélopartage et l'électrification des vélos permettraient selon [Zhang et al. \(2014\)](#) de favoriser davantage la mobilité liée à ce mode propre et ainsi encourager la multimodalité. Enfin les Pays-Bas sont allés encore plus loin en proposant une solution de vélo-train avec la possibilité d'emporter son vélo dans le train, ou bien de le garer à la gare et d'en prendre un autre par la suite, ou encore de tout simplement louer un directement dans la gare ([Eric J.E. Molin & al., 2010](#)).

L'amélioration de la multimodalité passe également une meilleure transition entre l'automobile et les transports en commun. A défaut de promouvoir la voiture individuelle, les services d'autopartage ou encore de taxi-ferroviaires sont aussi à privilégier dans le cadre du développement de la multimodalité selon [Eric J.E. Molin & al. \(2010\)](#). En effet, une solution de taxi-ferroviaires a été implémentée par les chemins de fer néerlandais pour faire concurrence au taxi classique grâce notamment à des prix attractifs, permettant de convertir des trajets monomodaux (taxi seul) en trajets multimodaux (taxi et train).

Pour [Polydoropoulou & Ben-AkivaA. \(2001\)](#), le métro est le concept préféré sur l'ensemble des offres de transports en commun, loin devant le d'autobus mêmes très modernes. Cependant, le coût élevé de la construction et de l'exploitation du métro signifie que dans la majorité des villes, les réseaux de bus étendus continueront à se développer ([Yuji Sh, 2023](#)). La voiture multimodale elle reste majoritairement utilisée pour atteindre les réseaux ferroviaires denses selon [Ralph Buehler & al. \(2007\)](#), comme ceux des centres-villes par exemple. Au contraire, les gares ferroviaires situées dans les zones à faible ou moyenne densité sont peu empruntées et la multimodalité y est plus faible ([Haochun & Yunyi, 2023](#)) que dans les grands centres urbains denses. Selon ces mêmes auteurs, l'augmentation de la connectivité entre les gares et les bus permet d'augmenter l'accessibilité et l'offre de déplacement, ce qui est bénéfique pour l'urbanisation des zones suburbaines marginalisées. Toujours selon [Haochun & Yunyi \(2023\)](#), le transport par autobus, avec son grand nombre, sa large distribution et sa relative flexibilité, reste l'un des meilleurs échanges avec le transport ferroviaire. Cette configuration optimisée favorise grandement la multimodalité des usagers.

Il est souvent question des réseaux de transport en commun ou des mobilités douces (marche et vélo) lorsque l'on parle de multimodalité mais peu d'attention est accordée aux alternatives récentes qui se développent dans les centres urbains denses notamment. Les services de micro mobilité partagée ont des effets positifs significatifs sur la réduction de la congestion pour [Xavier Bach & al. \(2023\)](#) et augmentent également la flexibilité des transports tout en contribuant à réduire l'encombrement des réseaux de transport public saturés. Ce type de transport inclut les vélos électriques en libre-service mais également les trottinettes électriques, segways ou même les scooters électriques. Ce système est généralement accompagné d'applications permettant de trouver rapidement un moyen de locomotion disponible à proximité et permet de faire rapidement le lien entre le domicile et le mode principal de transport par exemple (ex : le train ou le métro).

Toujours selon les auteurs, Barcelone serait la troisième ville comptant le plus grand nombre de permis de cyclomoteurs après Bangalore et Taipei qui sont alors utilisés comme véritables alternatives à la

voiture ou même les autres transports en communs. Les micro mobilités sont en général plus rapides que les autres modes et restent surtout utilisés à des fins de trajet en complément de l'offre existante (Felix Schwinger & al., 2022). On notera, toujours d'après les derniers auteurs, que les vélos électriques sont souvent privilégiés pour les longs trajets par rapport à l'utilisation des trottinettes électriques sans doute imputé à leur confort plus élevé, mais cela dépend également de l'offre disponible. Souvent, il y a davantage de trottinettes que de vélos mis à disposition. Enfin, selon Aoyong Li & al. (2022), les trottinettes électriques qui s'appuient sur l'avantage de la flexibilité et de la commodité pourraient mieux satisfaire les voyages pour les loisirs, les loisirs et les activités touristiques. Les auteurs indiquent également que les e-scooters deviennent de plus en plus populaires dans de nombreuses villes du monde (Aoyong Li & al., 2022).

On a parlé jusque-là des modes de transport en eux-mêmes mais la multimodalité passe également par une meilleure interaction entre tous les modes, ce qui implique de réfléchir à une échelle plus large, celle des réseaux de transport. Laura Alessandretti & al. (2022) cherchent à changer la vision que l'on a des réseaux de transport non pas comme la somme de différents réseaux indépendants mais comme un système multicouches. Selon les auteurs, l'ajout de couches de réseau dans un système de transport multicouche peut aider à réduire la congestion mais n'empêche pas la friction pour passer d'un mode à l'autre, en particulier lorsqu'il y a des retards.

Il semble, selon ces auteurs, plus pertinent de compléter les connexions intra-couche présentes dans le système en densifiant par exemple les réseaux de trottoirs piétonniers et de zones piétonnes pour que ce réseau soit équivalent à celui des routes vélo par exemple. De même, un réseau d'autobus est généralement plus dense et plus fréquenté qu'un réseau ferroviaire urbain qui dessert de plus grandes distances, et les micro mobilités permettraient de désengorger les réseaux de bus, notamment pour les derniers kilomètres.

La notion de hubs de transfert est utilisée par Yifan Yue & al. (2021) pour désigner des nœuds qui facilitent le transfert entre les lignes ou les modes de transport en commun. Une vision nodale des réseaux de transport, où chaque nœud correspond à une intersection entre les réseaux, permet d'équilibrer la capacité de l'ensemble du réseau multimodal et d'assurer l'équilibre de chaque sous-réseau de transport (bus, train, route, etc.). Cette approche permettrait selon les auteurs toujours fortement limiter la congestion routière et de favoriser grandement la mobilité durable. Enfin, les auteurs alertent sur le fait qu'une fois que la coordination de la capacité des nœuds dans d'autres sous-réseaux est dépassée, cela entraîne une défaillance en cascade qui a un impact sur l'ensemble du réseau de transport.

L'accès à l'information est primordial dans le cadre de la multimodalité car en étant mieux informés sur l'état des transports, les usagers auront plus tendance à les emprunter et à effectuer des combinaisons multimodales. Bai & Kattan (2014) défendent qu'avec la présence d'informations en temps réel encourage davantage les voyageurs de moins de 45 ans, les automobilistes et les usagers occasionnels du transport en commun à effectuer plus de trajets en transport en commun réel. D'autres auteurs, comme Nguyen Hoang-Tung & al. (2021) indiquent que le nombre de lignes de transports augmente à mesure que le nombre de passagers augmente grâce à l'introduction d'un système d'information en ligne donnant des indications sur l'état du trafic. A l'inverse, le manque d'information fiable entraîne une désorientation accrue des voyageurs qui ont alors plus tendance à faire de mauvais choix lorsqu'ils doivent sélectionner un mode de transport à prendre (Chorus et al., 2007), ce qui impacte directement leur multimodalité.

2.4 Facteurs liés au contexte

On a vu jusque-là tout un ensemble de facteurs objectifs liés à l'utilisateur même ou au transport en tant que tel. Voyons maintenant d'autres facteurs objectifs liés au contexte du voyage.

Le choix du mode de transport dépend grandement du moment de la journée pour [Eric J.E. Molin & al. \(2010\)](#), car les voyageurs sont moins enclins à voyager le soir et la nuit que de jour pour des raisons de sécurité et de commodité (en particulier les femmes). Ces mêmes auteurs indiquent par exemple que la probabilité de choisir les modes lents (vélo, marche), comme mode de sortie après le mode de transport principal (ex : le train), est augmentée lorsque l'on voyage à la lumière du jour. D'autres auteurs comme [Stephan Krygsman & al. \(2004\)](#) indiquent que le temps d'accès et de sortie aux transports augmentent durant les heures creuses mais le temps par rapport aux heures de pointe et cela s'explique par une offre moins importante en transports durant ces périodes.

Au contraire, selon [Bai & Kattan \(2014\)](#), la majorité des déplacements des navetteurs ont lieu pendant les périodes de pointe, lorsque la fréquence des transports en commun est assez élevée là où les déplacements en taxi ou en voiture se font en dehors des heures de pointe, du fait de la faible fréquence des transports et d'un réseau routier non congestionné. Selon ces derniers, les heures de pointe sont caractérisées par de meilleurs niveaux d'accessibilité des correspondances en moyenne, mais des performances sont moins bonnes sur les stations de métro à fort volume de correspondance par rapport aux heures creuses. La congestion durant les heures de pointe est donc un frein à la multimodalité.

La multimodalité peut être évaluée à l'échelle de la journée comme à l'échelle de la semaine. La multimodalité est ainsi plus élevée à l'échelle hebdomadaire qu'à l'échelle diurne pour [Ralph Buehler & al. \(2007\)](#). D'un autre côté, les usagers utilisant un service de train en ville plus d'une fois par semaine sont moins susceptibles de changer de moyen de transport en cas de retard sur la ligne, et auront plus tendance à attendre le prochain train ([Bai & Kattan, 2014](#)). En cas de retard, les usagers sont donc en général moins multimodaux car ils gardent leurs habitudes de transport, en particulier dans les transports en communs. L'habitude est un facteur tellement important la multimodalité des individus car plus le nombre augmente déplacements quotidiens augmente et plus la personne est susceptible d'être multimodale ([Blumenberg & Pierce, 2013](#)). [Eva Heinen & Kiron Chatterjee \(2015\)](#) ont noté qu'entre 2001 et 2009, la proportion d'automobilistes multimodaux hebdomadaires a augmenté.

Le temps d'attente est un facteur très important à considérer lorsque l'on établit une chaîne multimodale car, d'après [Chieh-Hua Wen & al. \(2012\)](#), la plupart des usagers y sont très sensibles et cela a un impact sur leurs choix lors du passage d'un mode à un autre. Ainsi, selon les auteurs, les voyageurs recherchent en priorité les transports en commun qui offrent des services à haute fréquence et sinon ils se tourneront vers les voitures, les taxis ou encore les navettes express. [Riccardo Gallotti & Marc Barthelemy \(2014\)](#) indiquent d'ailleurs que, dans le cadre d'un réseau urbain multimodal, les temps d'attente inter-couches et les temps de marche commencent à jouer un rôle significatif lors des longs trajets.

Au contraire, plus les couches ferroviaires et métropolitaines sont rapides, plus le trajet minimal a tendance à être multimodal d'après eux. Le retard est intimement lié au temps d'attente et selon [Bai & Kattan \(2014\)](#), un retard trop important pousse les usagers vers d'autres itinéraires de transport dans près de 50% des cas. Finalement, ces variations temporelles doivent être soigneusement prises en compte afin d'améliorer la connectivité entre les différents modes de transport public dans le but d'établir un système de transport en commun intégré efficace et efficient ([Yuji Sh, 2023](#)).

Un autre frein à la multimodalité est le coût des transports, un facteur tout aussi important (voir plus important) que la commodité même du système. [Chieh-Hua Wen & al. \(2012\)](#) indiquent par exemple que les voyageurs du TGV sont bien plus affectés par le coût du transport que le temps de trajet car ils préféreront souvent mettre plus de temps pour arriver à destination si le billet est moins cher. Dans une autre mesure, selon [Marco Diana \(2010\)](#), l'augmentation des frais de stationnement décourage de nombreux voyageurs de se rendre aux gares en voiture, surtout quand leur niveau de revenu est bas ou que le trajet est court, ce qui peut les inciter à davantage prendre les transports.

[Chieh-Hua Wen & al. \(2012\)](#) ajoutent qu'une part non négligeable des voyageurs préfèrent les modes de déplacement à bas prix tels que le train ou la moto au détriment du taxi par exemple. Augmenter les niveaux tarifaires tout en préservant la structure tarifaire est plus envisageable par les usagers que d'avoir un tarif unique mais qui favorise les petits transferts payants, les laissez-passer mensuels et les billets prépayés d'après [Marco Diana \(2010\)](#). Toujours selon l'auteur, contre toute attente, une réduction des tarifs peut également être avantageux pour les compagnies de transport car cela attire de nouveaux usagers et serait même plus bénéfique sur le long terme que de simplement augmenter les prix. On notera enfin que les modes les plus coûteux, qui impliquent à la fois des taxis et des Greenwheels ou encore des voitures, sont plus souvent choisis à des fins professionnelles qu'à des fins récréatives et cela au détriment du train ([Eric J.E. Molin & al., 2010](#)).

Dans une autre mesure, l'utilisation des transports publics dépend en grande partie de la commodité du système ([Stephan Krygsman & al., 2004](#), [Polydoropoulou & Ben-Akiva, 2001](#)), car une demande d'effort physique trop importante lors de l'accès aux transports dissuade les usagers d'utiliser ces transports. L'augmentation du temps d'accès (ou de sortie) aux transports tend également à dissuader les usagers de les utiliser et selon les auteurs, les voyageurs sont même prêts à accepter des temps de trajets plus longs tant que l'accès et la sortie sont plus rapides. La commodité des transports passe également par développement de points d'échange (ex : parkings relais, pôles de mobilité, etc.) d'après [Eric J.E. Molin & al. \(2010\)](#) afin d'accroître l'efficacité et la durabilité du système entier de transport. Au contraire, toujours selon les derniers auteurs, si les modes de sortie proposés dans les ensembles sont jugés comme insuffisants, les voyageurs peuvent alors décider d'utiliser la voiture plutôt qu'un autre mode de transport, voire ne pas voyager du tout. Ainsi pour [Chieh-Hua Wen & al. \(2012\)](#), l'amélioration d'un mode d'accès vers ou depuis une gare, par exemple, est susceptible d'encourager les voyageurs à passer d'autres modes d'accès et de sortie et même d'inciter les voyageurs d'autres modes principaux à utiliser les chemins de fer.

D'un autre côté, la disponibilité du stationnement est négativement associée au choix de la marche comme mode d'entrée ou de sortie ([Sungyop Kim & al., 2007](#)). Certaines mesures comme des voies réservées aux autobus réduisent le temps de circulation, mais d'autres comme les nouveaux autobus à plancher surbaissé réduisent le temps d'arrêt d'après [Ehab I. Diab & al. \(2015\)](#). Ainsi la recherche de multimodalité ne passe pas uniquement par des temps de trajet plus courts mais également par des temps de transferts inter-modes plus courts (augmentation de la commodité du réseau de transports). Les services de micro mobilité en libre-service (ex : les vélos électriques) peuvent également être intégrés à une offre de bus à haut niveau de service via un seul abonnement afin de faciliter le transfert d'un mode à l'autre ([Zhang et al., 2014](#)). Les micro-mobilité permettent de compléter efficacement les modes de transports principaux (bus, trains, métro) notamment pour effectuer les derniers kilomètres.

La multimodalité est étroitement liée à l'offre de transport en elle-même et donc la disponibilité des différents modes de transport. Bien que l'offre de transport varie à la fois spatialement et temporellement en fonction de la demande, la variation limitée de l'offre ne parvient pas à compenser totalement la variation beaucoup plus importante de la demande selon [Yuji Sh \(2023\)](#). Ainsi, de nombreuses stations de transports en commun sont sous-utilisées quand d'autres n'arrivent pas à faire

face à la demande. [Polydoropoulou & Ben-AkivaA. \(2001\)](#) ajoutent que la disponibilité des modes de transport en commun dépend également de la façon dont ils ont été intégrés dans le réseau global de transports, car le fonctionnement indépendant des lignes de transport nuit considérablement à l'efficacité de l'ensemble du système, et donc réduit la commodité globale (diminution de la multimodalité).

La commodité des transports s'exprime également par un accès plus ou moins facilité aux transports. [Sungyop Kim & al. \(2007\)](#) indiquent que la distance entre le domicile et la gare est d'office négativement associée à la marche (plutôt associée à la voiture) tout comme la présence de zone de park-and-ride dans les zones suburbaines. L'utilisation des transports en commun implique généralement un déplacement multimodal d'après [Blumenberg & Pierce \(2013\)](#) (ex : marcher jusqu'à l'arrêt de bus, d'utiliser l'autobus, puis de marcher jusqu'à la destination finale) et les personnes qui n'ont pas d'autre choix que de prendre le bus marchent plus longtemps que celles qui ont une autre solution, conventionnels, ce qui réduit le nombre de longs trajets accessibles à pied ([Yang Jiang & al., 2012](#)). [Sungyop Kim & al. \(2007\)](#) ajoutent qu'une disponibilité accrue en parkings (notamment des parkings relais) permet d'attirer plus d'usagers du train en offrant des stationnements incitatifs, en particulier quand on sait que le stationnement devient plus rare et plus cher dans les centres urbains denses.

Dans une autre mesure, les services de taxi partagé sont très appréciés par les passagers des autobus et des trains selon [Chieh-Hua Wen & al. \(2012\)](#), ce qui tend à diminuer le nombre de voitures sur les routes et limite ainsi la congestion dans les villes. Certains auteurs comme [Stephan Krygsman & al. \(2004\)](#) notent que, pour les usagers de bus, de tram ou de métro, il n'y a pratiquement pas de différence entre l'accès et la sortie notamment car le taux d'interconnectivité est plutôt modeste. Un meilleur aménagement urbain autour des gares et arrêts de bus incite plus d'usagers à utiliser les services de BHNS (bus rapide) ou de train ([Yang Jiang & al., 2012](#)) et augmente ainsi la commodité globale du système de transport. Enfin la commodité du système se retrouve également au travers des micro-mobilité (ex : les vélos électriques en libre-service) qui facilitent grandement la transition vers les modes principaux de transports en commun (ex : bus, train, métro) pour [Zhang et al. \(2014\)](#) ce qui tend là encore à encourager la multimodalité.

La possession d'un abonnement de transport augmente la multimodalité selon [Eva Heinen & Kiron Chatterjee \(2015\)](#). De nombreux abonnements permettent aujourd'hui d'avoir accès à de nombreux sous-réseaux de transports très différents et ce en même temps, notamment dans les grandes villes (ex : pass Navigo à Paris). Ces pass tout compris limitent le temps perdu au guichet pour acheter un ticket et assurent une certaine fluidité de déplacement au sein d'un réseau de transport global optimisé. Le temps perdu dans les réseaux de transport vient généralement de l'indépendance des sous-réseaux (souvent entretenus par des gestionnaires différents) ce qui entraîne une friction entre ces réseaux, conduisant à des retards. La mise en place d'une offre unifiée permet de pallier ces problèmes, ce qui se matérialise généralement par un abonnement global, très pratique et apprécié par les usagers (ex : le pass Navigo en région parisienne). Cela encourage l'utilisation des transports en communs notamment et donc favorise la multimodalité.

La multimodalité est également étroitement liée à la notion de confort car plus le mode de transport est agréable et/ou accueillant et plus les personnes auront envie de l'utiliser. Parmi toutes les questions liées au confort, la qualité de service est primordiale selon [Yang Jiang & al. \(2012\)](#). Cela inclut notamment la possibilité d'acheter un billet avant l'embarquement (ex : en ligne ou en gare), des mesures de contrôle et de gestion des flux (notamment piétons), une fréquence plus importante des transport, ou encore la prise en charge des bagages (entre autres). Car pour [Chieh-Hua Wen & al.](#)

(2012), étant donné que les modes d'accès privés ont des parts de marché élevées, les modes d'accès publics doivent offrir une qualité de service élevée pour attirer les utilisateurs des modes privés.

Cela va bien au-delà du transport en lui-même ou de la gare car certaines compagnies incluent dans l'offre de transport un transfert gratuit vers un autre mode de transport (bus, services de micro-mobilité, taxi, etc.) ou encore mettent en place des parkings relais dans les gares avec généralement une restriction du stationnement automobile à proximité (Sungyop Kim & al., 2007). La possibilité de trouver un siège est également une notion importante quand on parle de confort d'après Polydoropoulou & Ben-AkivaA. (2001), notamment chez les personnes qui se fatiguent vite (ex : les personnes âgées, femmes enceintes) et cela est particulièrement le cas dans les transports en communs classiques ou une place n'est pas attribuée (train de banlieue, bus, métro, tram, etc.). En lien avec la notion de confort, les modes lents de marche et de vélo sont privilégiés lorsque l'itinéraire du trajet est connu car la charge mentale associée au trajet est considérée comme faible chez les usagers (Eric J.E. Molin & al., 2010). A l'inverse, les modes qui incluent un chauffeur (ex : taxi, voiture) ou un véhicule personnel (ex : Greenwheels, voiture personnelle) sont privilégiés si l'on ne sait pas comment atteindre la destination.

La fiabilité et la sécurité sont aussi des facteurs importants à considérer lorsque l'on parle de multimodalité. Ehab I. Diab & al. (2015), indiquent ces deux facteurs constituent même le fondement de la satisfaction des voyageurs. Ces mêmes auteurs indiquent que la fiabilité est importante pour les opérateurs des transports même car une meilleure fiabilité permet de réduire les coûts d'exploitation et donc d'améliorer les bénéfices. Mais dans le même temps, cela permet aux usagers de payer moins cher leur titre de transport, une situation bénéficie donc à tout le monde. Pour Duthilleul et al. (1999), les compagnies de transport doivent rivaliser avec l'automobile, solution toujours très attractives, et elles se doivent donc de proposer une offre de qualité et à des prix adaptés.

La peur de la criminalité ou même le simple niveau de criminalité perçu décourage l'utilisation des moyens de transit (ex : bus) chez les individus selon Sungyop Kim & al. (2007). Les femmes ont plus tendance à éviter de marcher ou d'attendre dans les gares la nuit et/ou dans les gares où la criminalité est plus signalée et celles-ci privilégieront toujours la voiture aux autres modes (Sungyop Kim & al., 2007, Eric J.E. Molin & al., 2010). Ces auteurs affirment que l'augmentation des feux et des patrouilles de police ou de sécurité peut améliorer la sécurité et aider les femmes à se sentir plus à l'aise, voire les encourager à utiliser davantage les transports. Et cela peut également améliorer la mobilité des femmes qui ont moins d'accès à l'automobile personnelle, et donc favoriser leur multimodalité.

Quand on parle de sécurité, cela fait aussi référence à la notion d'accident et fait écho à ce qui a été dit sur la fiabilité et l'accessibilité. Les auteurs Yang Jiang & al. (2012) et Sungyop Kim & al. (2007) indiquent que la densité des trottoirs est positivement associée au pourcentage de marche pour se rendre au transport en commun. La présence de marche augmente le risque de trébucher et donc de se blesser, en particulier lorsqu'il y a du monde ou que l'on est pressé. La présence de marches réduit la probabilité d'accès à pied aux gares par exemple d'après les auteurs, ce qui a de fait un impact négatif sur la multimodalité des usagers. D'autres auteurs comme Zhang et al. (2014) alertent sur le risque d'accident lorsque l'on traverse une route à pied ou à vélo, ce qui constitue le principal facteur de dissuasion à l'égard du cyclisme.

Un autre facteur lié à la multimodalité est le motif du déplacement qui, comme le décrivent Blumenberg & Pierce (2013), permet de prédire la distance parcourue par un déplacement multimodal. Stephan Krygsman & al. (2004) ont identifié que le temps de trajet ainsi que le temps d'accès et de sortie aux transports diminuent lorsque les voyageurs se déplacent pour se rendre à l'école, aux loisirs ou encore pour faire des achats. D'autres auteurs comme Eric J.E. Molin & al. (2010) et Chieh-Hua Wen

& al. (2012) indiquent également que les modes moins coûteux (et souvent plus lents) sont d'ailleurs plus préférables à des fins récréatives qu'à des fins commerciales.

La voiture, les taxi et greenwheels sont au contraire des modes privilégiés dans le cadre de voyages professionnels (Eric J.E. Molin & al., 2010). Ces mêmes auteurs montrent que les détenteurs d'un abonnement d'autobus sont plus enclins à voyager à des fins récréatives. Le vélo est le mode de déplacement privilégié pour les trajets domicile-travail dans les grands centres urbains et souvent devant la marche d'après Zhang et al. (2014), car il demande notamment moins d'efforts physiques et est plus rapide (en particulier dans le cas de vélos électriques). Cependant ces mêmes auteurs indiquent que ce mode reste dangereux, notamment vis-à-vis du trafic, et sera donc dans de nombreux cas abandonné au profit de la voiture notamment.

La multimodalité dépend également de la configuration urbaine, tant en termes de densité que d'adaptation au terrain (topographie notamment). Selon Stephan Krygsman & al. (2004), les modes doux comme la marche ou le vélo sont très sensibles à la densité urbaine, plus la densité urbaine est importante et plus la multimodalité sera favorisée. En effet d'après les auteurs, l'augmentation de la densité entraîne généralement une réduction du temps de trajet, sauf dans le cas où tout le monde converge vers un seul point (ex : une seule station).

Dans ce cas, des densités plus élevées peuvent conduire à des distances nettement plus courtes entre la gare et les destinations finales ce qui augmente la congestion et donc réduit la multimodalité. D'un autre côté, selon Blumenberg & Pierce (2013), les voyageurs unimodaux sont plus susceptibles de vivre dans des quartiers à faible densité, là où généralement l'offre de transports est plus faible du fait du faible nombre d'habitants (demande potentielle en transport faible). Au contraire, les ménages sans automobile ont tendance à déménager dans des quartiers urbains denses des centres-villes offrant de nombreuses options de modes de transport (Blumenberg & Pierce, 2013).

D'après Sungyop Kim & al. (2007) une utilisation importante des terres, liée à une concentration des activités, augmente la propension à marcher jusqu'aux stations de transport en commun là où les zones d'utilisation mixte des terres sont associées négativement au choix du mode de la marche. Ainsi les zones industrielles et commerciales ou encore de bureaux sont liées à une plus grande propension aux modes d'autobus et de marche selon ces mêmes auteurs. La topographie est généralement moins propice au vélo selon Zhang et al. (2014), notamment dans les villes situées dans des zones montagneuses et vallonnées. D'après le travail de Xavier Bach & al. (2023), les pentes supérieures à 4 % rendent vraiment difficile le vélo comme moyen de transport quotidien. Ils affirment également que la densité et la zone de couverture géographique d'un service de micro mobilité partagée jouent un rôle important dans son succès.

Pour parler plus spécifiquement de la densité de la population liée à la forme urbaine, les auteurs Stephan Krygsman & al. (2004) ou encore Marco Diana (2012) indiquent que le temps d'accès et de sortie des transports est plus important au cœur des villes métropolitaines qu'entre la ville et l'extérieur (particulièrement plus faible dans les villes de plus de 50 000 habitants, dans les petites villes et banlieues des villes métropolitaines). Cependant d'après, Yuji Sh (2023) dans le cas de grandes métropoles comme Shanghai, l'essentiel des transferts se déroulent à l'extérieur du centre-ville et généralement pendant les heures creuses de fin de soirée début de nuit.

Ralph Buehler & al. (2007) notent également que plus la densité de population est importante et plus grande est la probabilité d'être un automobiliste multimodal et cela est aussi renforcé par la proximité à un système ferroviaire. D'ailleurs, les zones à faible densité encouragent la construction de train rapide en prévision de l'augmentation de la population et de l'emploi à l'avenir selon Sungyop Kim & al. (2007). Marco Diana (2012) note cependant que la proportion de travailleurs n'est pas radicalement

différent selon si on habite en centre-ville ou en périphérie mais les personnes à faible revenu résidant dans les quartiers denses du centre-ville, comptent sur une offre de transport plus importante pour pouvoir se déplacer (Blumenberg & Pierce, 2013).

La météo influence aussi grandement les choix des usagers en termes de transports. Pour Eric J.E. Molin & al. (2010), les voyageurs se font plus rares par temps de pluie, et les alternatives à la marche et au vélo seront préférées, notamment si elles offrent un abri lors de l'attente (abris bus, gare par exemple). Au contraire, les modes lents sont plus souvent utilisés par beau temps et ce tout autant que l'automobile ou le taxi. Le choix multimodal dépend aussi du climat (ou de la saison) car, d'après Bai & Kattan (2014), les voyageurs sont moins susceptibles de prendre une voiture, un taxi ou un autre moyen de transport en commun au cours d'une journée d'été typique où il fait très chaud. De même, pour les mêmes auteurs, les voyageurs ont plus tendance à attendre le prochain train par temps chaud. Mais dans le même temps, certains transports en commun ou modes doux qui demandent d'être en contact avec l'extérieur sont moins privilégiés par temps froid, notamment au cœur de l'hiver (Zhang et al., 2014).

Le choix du mode de transport dépend également de la quantité de bagages embarquée pour Eric J.E. Molin & al. (2010), car les modes motorisés, en particulier les alternatives de taxi, sont préférés si des bagages lourds doivent être transportés, au détriment du train par exemple. De plus, la probabilité de choisir les modes lents comme vélo et marche comme mode de sortie est augmentée lorsqu'il n'y a pas de bagages lourds à porter d'après ces mêmes auteurs. Enfin, les modes motorisés sont privilégiés lorsque des animaux de compagnie sont présents d'après Ralph Buehler & al. (2007). Dans la même veine, voyager avec d'autres personnes rend l'alternative du taxi moins chère et peut rendre certaines options de vélo plus problématiques, car tous les voyageurs du groupe auraient besoin d'avoir un vélo (Eric J.E. Molin & al., 2010). Voyager en compagnie augmente également le choix des modes de transport autres que le train, augmentant ainsi la probabilité que la voiture soit choisie, selon ces derniers auteurs.

Les pouvoirs publics jouent un rôle majeur sur la multimodalité car des politiques qui encouragent l'utilisation de la marche, du vélo et des transports en commun pour tous contribuent à accroître l'utilisation de la voiture multimodale chez les automobilistes monomodaux d'après Ralph Buehler & al. (2007). Cela permet également de renforcer la tendance à une part plus élevée de la multimodalité chez les automobilistes qui utilisent déjà des alternatives à la voiture mais qui souhaitent augmenter l'intensité de leur multimodalité selon les auteurs. De même, les politiques qui facilitent l'accès aux transports publics et qui augmentent la densité de population avec des distances de trajet plus courtes peuvent contribuer à accroître encore la part des conducteurs multimodaux.

2.5 Facteurs psychologiques

Nous avons traité jusque-là des facteurs objectifs qui impactent la multimodalité des usagers. Ces facteurs sont généralement très bien documentés par la littérature mais peu d'études portent sur des facteurs plus subjectifs liés notamment au comportement des usagers et de la perception que ces derniers ont des transports.

Nous avons précédemment vu que la notion de confort dans les transports avait un impact direct sur la propension des usagers à utiliser ces transports. Cette notion est étroitement liée à la notion de perception du service passe par différents paramètres comme on l'a vu : la fiabilité, la commodité et le confort entre autres. D'après Marco Diana (2010), la perception des transports est influencée par des

processus purement rationnels (ex : décisions de changement de mode suite à un retard) mais également par d'autres biais psychologiques moins évidents comme des motivations sociales spécifiques, la personnalité des usagers, des dilemmes sociaux (mécanismes psychologiques) ou encore par des processus de dissonance cognitive.

La perception du service diffère aussi selon le type de multimodalité car, d'après [Thomas Klinger \(2017\)](#), les cyclistes multimodaux ont une attitude plus positive à l'égard des transports publics que les autres groupes (cyclistes classiques, piétons multimodaux, automobilistes multimodaux...). [Zihao An & al. \(2021\)](#) affirment également que l'âge influence la perception au service car les différences générationnelles dans le domaine de l'automobile (et donc de la multimodalité) pourraient être principalement attribuées à la dégradation des conditions de transport au cours des dernières décennies.

De même, d'après [Marco Diana \(2012\)](#), les retraités multimodaux vivant dans les centres-villes privilégient plus le confort, tandis que les personnes qui prennent le métro préfèrent la ponctualité. La perception des transports n'est pas la même selon la configuration urbaine car, selon cette même autrice, on retrouve un niveau plus élevé de satisfaction vis-à-vis des transports en commun dans les petites villes et alors qu'il est plus faible dans les villes métropolitaines (sans distinction significative entre le centre-ville et les banlieues d'ailleurs). Il faut cependant noter que les transports en commun doivent être bien évalués par les automobilistes dans les petites villes avant d'être considérés comme une véritable alternative à la voiture.

Toujours selon [Marco Diana \(2012\)](#) ou encore [Bai & Kattan \(2014\)](#), les voyageurs qui perçoivent mal l'information en temps réel ont tendance à utiliser un autre moyen de transport pendant leurs trajets domicile-travail. Les personnes moins familières avec un mode sont d'ailleurs davantage influencées dans leurs choix de transport par des évaluations plus subjectives (le degré de satisfaction lié au service), d'après ces mêmes auteurs. A leur décharge, il est affirmé par [Chorus et al. \(2007\)](#) les réseaux de transport urbain multimodal deviennent de plus en plus complexes, ce qui les rend du vélo étant pour les usagers utilisant peu fréquemment les transports en communs (comme les automobilistes par exemple).

D'autres modes de transport comme la marche ou le vélo, qui ne sont pourtant pas si faciles à utiliser lorsque l'on vit en banlieue, peuvent être même envisagés si les transports en commun sont trop mal évalués ([Marco Diana, 2012](#)). Reste que pour [Chorus et al. \(2007\)](#), les voyageurs n'ont pas une connaissance précise sur les attributs liés à leur voyage (temps de trajet, coûts, etc.) prennent globalement de moins bonnes décisions dans leurs choix de transports et sont par conséquent sont globalement moins satisfaits (ex : associés à une perte de temps comme le soulignent [Eva Heinen & Kiron Chatterjee \(2015\)](#)).

Pour [Nguyen Hoang-Tung & al. \(2021\)](#), la satisfaction d'un transport donné peut être impactée par la création d'embouteillage par un autre mode de transport (ex : une voie exclusive BHNS introduite sur voie à fort trafic). C'est en particulier pour cela que les bus sont généralement perçus comme un mode de transport de moindre qualité en plus du manque de confort (sièges vétustes), ce qui peut avoir un impact négatif sur la motivation à introduire une ligne de bus rapide (BHNS) de haute qualité selon les mêmes auteurs. L'augmentation des tarifs des billets a également un impact négatif sur l'accessibilité des transports et donc sur la satisfaction associée, notamment pour les courts trajets ([Marco Diana, 2010](#)).

Selon [Zihao An & al. \(2022\)](#) ou encore [Duthilleul et al. \(1999\)](#), les personnes ayant une attitude négative à l'égard de la marche (ex : à cause de l'effort physique associé) sont particulièrement insatisfait lorsque le manque d'indications et de signalétique liées aux transport les contraignent à marcher plus que prévu.

A l'inverse, les passagers qui choisissent de marcher davantage dans le but d'utiliser le service le plus rapide, sont plus sensibles à l'attente et au gain de temps sur l'ensemble du trajet (Ehab I. Diab & Ahmed M. El-Geneidy, 2014). Yongsung Lee et al. (2020) ajoutent d'ailleurs que les voyageurs n'ayant pas pour habitude d'utiliser la voiture mais plutôt les modes doux (marche, vélo), n'ont pas l'impression d'être limités dans l'horaire des déplacements ou dans le choix des modes de déplacement et ont ainsi une attitude plus positive à l'égard de ces modes que les autres.

Par ailleurs, les auteurs William Jen & Kai-chieh Hu (2003) indiquent que la valeur perçue a un impact plus important que l'attractivité des modes alternatifs, ce qui indique que le désir des passagers de racheter le service est influencé par la valeur du service offert. De plus, toujours selon les mêmes auteurs, dans le cas où le temps n'est pas pris en compte, les clients peuvent choisir d'autres compagnies d'autobus ou d'autres modes de transport, qui offrent une qualité de service similaire ou supérieure, mais des coûts perçus inférieurs.

On ajoutera que les achats deviennent même plus probables lorsque les avantages générés dépassent les coûts d'après, parmi lesquels : une interaction avec les passagers, des équipements assurant le confort des passagers (ex : sièges confortables), une commodité de service (accessibilité), un système d'information clair ou encore un soutien à la gestion des opérations (William Jen & Kai-chieh Hu, 2003). Ehab I. Diab & Ahmed M. El-Geneidy (2014) ajoutent enfin que la qualité de service perçue par les utilisateurs est positivement liée à la satisfaction et est considérée comme le principal facteur de fidélisation et de comportement des usagers des transports en commun (ex : le bus).

La multimodalité varie aussi selon la culture d'un pays ou d'une région donnée car par exemple, les personnes âgées sont plus multimodales en Allemagne qu'aux Etats-Unis d'après Heinen & Mattioli (2019), notamment parce que le système de transport en Allemagne soutient mieux les personnes âgées utilisant différentes options de mode (Eva Heinen & Kiron Chatterjee, 2015). De même, les personnes qui se déplacent entre des villes représentant des cultures de mobilité différentes sont susceptibles de changer leur utilisation et leur combinaison de modes de déplacement à l'arrivée selon Thomas Klinger (2017). De fait, les personnes ayant déménagé dans une ville accueillante pour les transports en commun ou le vélo (ex : Amsterdam) sont plus susceptibles de devenir multimodaux que ceux qui ont déménagé dans des villes orientées vers l'automobile selon le même auteur. Les Néerlandais utilisent moins la voiture et les transports publics intra-urbains mais qu'ils font plus de vélo que leurs homologues dans d'autres pays européens et sont donc bien plus multimodaux.

Pour Yang Jiang & al. (2012), la vitesse moyenne des piétons varie d'un pays à l'autre car elle est inférieure dans les pays en développement d'Asie par rapport à l'Occident, ce qui a un impact direct sur l'utilisation de l'espace de la rue, et donc de la multimodalité (Yang Jiang & al., 2012). Le vélo électrique connaît un grand succès en Chine selon Zhang et al. (2014) car il permet de fortement limiter le trafic routier et donc la pollution, ce pourquoi il a été fortement encouragé par le gouvernement chinois dans les grandes villes. La culture du vélo était déjà présente dès les années 1960 en Chine mais ce mode de transport était alors réservé aux populations pauvres. En Europe au contraire ce mode est privilégié pour le gain de temps qu'il procure dans les centres urbains denses ou pour des raisons de santé (pratique du sport).

Les stéréotypes vis-à-vis de nouveaux modes comme les micromobilités varient également d'une région à l'autre d'après Zihao An & al. (2023) cela affecte directement leur déploiement. Les habitants des pays nordiques en Europe ont ainsi une perception bien plus positive et moins stéréotypée sur les micromobilités (e-scooters, vélo ou trottinettes électriques, etc.) que leurs homologues américains par exemple (trop peu de sécurité selon eux).

Au-delà de la culture, l'évolution des modes de vie jouent également un rôle important dans un contexte de prise de décision où l'information selon [Marco Diana \(2010\)](#) ou encore [Blumenberg & Pierce \(2013\)](#), ce qui a aussi un impact sur la multimodalité. Les modes de vie évoluent avec le temps et les auteurs comme [Zihao An & al. \(2021\)](#) remarquent un léger déplacement de la part modale de la voiture vers les transports publics depuis les années 1990, traduisant l'augmentation de la multimodalité à l'échelle mondiale. En effet, après la seconde guerre mondiale et dans un contexte de croissance économique, les baby-boomers occidentaux ont privilégié l'achat et la conduite de voitures privées au détriment des transports en commun. [Zihao An & al. \(2021\)](#) et [Ralph Buehler & al. \(2007\)](#) indiquent également qu'aujourd'hui, la réduction de la mobilité quotidienne des jeunes est imputée à des études plus longues par rapport aux précédentes générations et à une virtualisation accrue de la société (notamment avec le développement d'internet).

Les milléniaux choisissent souvent la multimodalité des voyages, au moins en partie à cause des effets de la crise économique (2008), qui a touché plus durement les jeunes adultes que leurs homologues plus âgés ([Yongsung Lee et al., 2020](#)). Des événements particuliers de la vie (vie indépendante des parents, le mariage et la maternité) sont globalement retardés par rapport aux générations précédentes, tandis que la poursuite d'études supérieures a augmenté et cela joue un rôle important sur la multimodalité des jeunes adultes qui sont moins enclins à posséder une voiture par exemple. A cela s'ajoutent des difficultés à acheter sa propre maison individuelle qui favorise au contraire les logements plus petits dans des quartiers urbains denses heureusement bien desservis en transports en communs.

D'après [Thomas Klinger \(2017\)](#), la numérisation croissante de la mobilité urbaine facilite de nouveaux comportements de déplacement, y compris la combinaison plus flexible et spontanée de différents modes de transport. Ainsi, les automobilistes multimodaux ont tendance à être aujourd'hui plus jeunes, à faire des études supérieures à l'école secondaire, à vivre dans des ménages sans voiture et à vivre dans des quartiers à forte densité avec accès à un système ferroviaire ([Ralph Buehler & al., 2007](#), [Heinen & Mattioli, 2019](#)). De plus, toujours d'après les auteurs, la possession d'une voiture est jugée comme moins abordable et moins nécessaire dans le cas de contextes familiaux défavorables, d'emploi plus précaires, au report des événements de la vie (ex : formation d'une famille, premier emploi permanent), de congestion du trafic ou plus généralement d'un contexte économique défavorable et la hausse des prix de l'essence.

A cela s'ajoutent des infrastructures plus nombreuses et de meilleure qualité pour la marche, le vélo et les transports en commun qui, comme le soulignent [Ralph Buehler & al. \(2007\)](#), encouragent également la multimodalité. Il faut cependant nuancer car selon [Heinen & Mattioli \(2019\)](#), la multimodalité a également diminué depuis les années alors qu'elle était en augmentation avant cela et alors que la part modale de la voiture restait stable dans le temps.

Au-delà de la culture et des modes de vie, la multimodalité est affectée par des phénomènes de modes ou des changements d'attitudes imputés à des événements plus ponctuels qui ont une influence sur les comportements des usagers. Les habitudes multimodales et les attitudes cognitives jouent un rôle important dans un contexte de prise de décision où l'information sur les nouvelles alternatives est incomplète comme l'indique [Marco Diana \(2010\)](#). L'auteur indique que même si les nouvelles technologies et nouveaux modes de transports (covoiturage, autopartage, etc.) peuvent améliorer les modes de transport existants, les habitudes conservent un rôle prépondérant dans les choix que l'on fait.

Ainsi, selon [Wokje Abrahamse & al. \(2009\)](#), l'utilisation de la voiture pour se rendre au travail dépend des normes personnelles de chacun (moralité) et également de la pression sociale ou la peur du regard

des autres (conformisme, mimétisme). Ces influences sont particulièrement fortes sur les déplacements urbains et sur de courtes distances que pour les déplacements longue distance pour [Marco Diana \(2012\)](#).

De plus, la volonté de réduire l'utilisation de la voiture pour les déplacements domicile-travail est corrélée à une obligation morale de réduire son utilisation car, pour [Wokje Abrahamse & al. \(2009\)](#), plus le conducteur se sent responsable des problèmes liés à ce mode (ex : pollution) et plus ils se sentent moralement obligés de réduire son utilisation. [Thomas Klinger \(2017\)](#) ajoute qu'à l'inverse, plus l'attitude à l'égard de l'utilisation de la voiture est positive et plus les personnes se rendent fréquemment au travail en voiture. De plus, les personnes qui possèdent un vélo et/ou un abonnement pour les transports publics ont tendance à combiner plus souvent les modes de transport ([Thomas Klinger, 2017](#)). L'environnement a un également un impact sur la multimodalité, car, toujours selon le même auteur, les personnes qui préfèrent le mode multimodal sont plus susceptibles de choisir un lieu de résidence qui facilite ce comportement (ex : proche d'un hub de transport ou en centre-ville dense). De même, si le voyageur voyage pour le plaisir, il a plus tendance à utiliser le bus ou à louer un vélo selon [Eric J.E. Molin & al. \(2010\)](#).

Selon [Marco Diana \(2010\)](#), le marketing autour des transports multimodaux ne devrait pas reposer sur la mise en avant de leur performance seule mais aussi sur leur fonctionnement afin de réduire la charge cognitive des clients potentiels et ainsi accroître la fréquentation des nouveaux systèmes. Les usagers conservent généralement leurs habitudes, car d'après [Bai & Kattan \(2014\)](#), les voyageurs dont le mode de déplacement principal est le vélo ou la marche sont moins susceptibles de choisir l'option de changement pour l'automobile ou le taxi de façon spontanée. De même, les utilisateurs fréquents du TLR préfèrent poursuivre leur trajet en TLR même lors de retard car c'est un mode de transport plus fiable que les autres et l'environnement d'attente est plus agréable que celui des bus.

Au contraire, toujours d'après [Bai & Kattan \(2014\)](#), les usagers dont le mode principal est l'automobile ont plus facilement tendance à prendre la voiture ou un taxi en cas de retard supérieur à 10 minutes sur la ligne de train. Reste que, plus un voyageur connaît d'alternatives, notamment via un service d'information en temps réel performant, plus il sera amené à changer ses habitudes vis-à-vis des transports et plus il sera multimodal ([Yongsung Lee et al., 2020](#)).

La Dissonance cognitive semble également jouer un rôle sur la multimodalité des usagers. Certains auteurs comme [Yang Jiang & al. \(2012\)](#) différencient le temps réel et le temps perçu lors d'un trajet, quand d'autres, comme [Clauss & Döppe \(2016\)](#), parlent d'une dimension affective liées au voyage. Selon les auteurs, les voyageurs sont de plus en plus enclins à combiner plusieurs modes de transport mais cela induit une source de stress qui biaise leurs choix, notamment lorsqu'ils rencontrent des problèmes aux interfaces entre les modes de déplacement. D'après, [Clauss & Döppe \(2016\)](#), dans ce cas, les options monomodales seront privilégiées pour diminuer le stress et dans une optique de recherche de l'efficacité (trajet le plus direct possible sans changements).

Un autre aspect important dans le choix du transport est le coût, on l'a vu, mais les auteurs montrent que les voyageurs ont tendance à réduire leurs frais de déplacement quel que soit leur niveau de revenu car le coût des billets ou des parkings relais est perçu plus négativement que celui du carburant pour la voiture. Toujours d'après [Clauss & Döppe \(2016\)](#), cela explique les intentions limitées de basculement vers les transports publics en particulier chez les automobilistes, même s'ils sont déjà multimodaux.

Le temps est également un paramètre important dans les choix modaux car les voyageurs y sont très sensibles, notamment les automobilistes, qui se montrent très réticents à l'idée de modifier leur choix de mode pour se tourner par exemple vers les transports en commun (jugés parfois comme moins

fiables). Ainsi, la flexibilité est un facteur clé qui peut pousser certains usagers à conserver leur voiture plutôt que de prendre d'autres modes de transport.

La perception des voyageurs à l'égard des modes de déplacement passe également par le confort et le sentiment de sécurité. Toujours selon [Clauss & Döppe \(2016\)](#), l'intérieur de la voiture est considéré comme un espace sûr et privé où l'utilisateur contrôle son environnement ce qui l'associe donc à la détente. La voiture privée est également associée à la liberté de se déplacer de façon autonome sans restriction par opposition à d'autres modes plus linéaires (ex : les trains) ou qui accumulent les retards échappant au contrôle individuel (ensemble des transports en communs), même les voitures peuvent se retrouver sur des routes congestionnées.

Ainsi, d'après les auteurs, les personnes ayant accès à la voiture bénéficient de plus d'avantages affectifs que les usagers des transports publics et le prestige et l'identification personnelle associés à ce mode expliquent pourquoi il est encore souvent privilégié par rapport à d'autres modes. A cela s'ajoutent une pression sociale qui tend à favoriser l'utilisation de la voiture par rapport aux autres modes même si cela évolue depuis récemment ([Clauss & Döppe, 2016](#)).

Certains auteurs comme [Marco Diana \(2010\)](#) ou [Blumenberg & Pierce \(2013\)](#) indiquent également que les traits de personnalité ou les préférences personnelles semblent jouer un rôle encore plus important dans un contexte de prise de décision où l'information sur les nouvelles alternatives de transport. En revanche, peu d'étude traite de la question dans le détail, mais les auteurs sont convaincus que cela joue sur la multimodalité de tout un chacun.

3- ANALYSE

3.1 Source de Données

3.1.1 NHTS 2022, présentation

Le but de cette étude est de savoir si la population américaine est multimodale d'une part et d'autre part de savoir quels sont les facteurs qui ont une influence sur la multimodalité dans les transports. Pour cela, une analyse statistique sera conduite sur l'Enquête nationale sur les déplacements des ménages (*National Household Travel Survey* ou *NHTS*), réalisée par le département des Transports des États-Unis (USDOT) sur l'année 2022 (précisément de janvier 2022 à janvier 2023 inclus).

Depuis 1969, le département américain conduit des études de ce type tous les 5 à 8 ans afin de saisir le mode de déplacement et le but du voyage des ménages américains, pour les déplacements quotidiens locaux et longue distance (tous modes confondus). L'objectif étant de comprendre les changements de comportement de voyage parallèlement à la composition démographique, économique et culturelle du pays. A titre d'exemple, le département des transports cherche à évaluer l'impact des nouveaux modes de transport comme les vélos en libre-service ou encore les e-scooters.

La présente étude comptabilise 31073 déplacements effectués par 10591 américains âgés de 5 à 92 ans, répartis sur l'ensemble du territoire des États-Unis. Est considéré ici comme déplacement le fait d'utiliser un mode de transport pour se rendre d'un point A à un point B sans coupure. L'étude visait à interroger des personnes utilisant au moins un mode de transport durant la journée afin de recueillir des données relatives au mode de déplacement ainsi que des données sur les usagers. L'étude a été

effectuée sur un échantillon volontairement très large et très diversifié afin d’obtenir les résultats les plus objectifs possibles et représentatifs de l’ensemble de la population américaine.

L’enquête de 2022 a été conçue pour suivre et suivre les tendances des principaux paramètres de comportement en matière de déplacement (ex : kilomètres parcourus par des personnes, le mode, l’objectif, l’heure de la journée de déplacement, durée du trajet, etc). L’étude est d’ailleurs la première de la série à considérer les longues distances (supérieur à 50 milles). Elle collecte également les données démographiques clés suivantes :

- Caractéristiques du ménage (ex : taille, possession d’un véhicule, type de logement, revenu et composition).
- Caractéristiques de la personne (ex : âge, sexe, race, origine ethnique, situation d’emploi et statut d’étudiant).
- Caractéristiques du véhicule (par exemple, année, marque, modèle et type de carburant).
- Plus de détails sur les élèves (une variable spécifique décrivant le statut de l’élève, le niveau actuel, l’emplacement de l’école et le mode typique d’école).
- Des questions plus précises sur emploi, en particulier celles relatives au télétravail, à l’utilisation des véhicules personnels à des fins commerciales et au mode de travail typique.
- Les détails de voyage pour tous les déplacements effectués par les membres du ménage âgés de 5 ans et plus pendant une période désignée de 24 heures.

Le NHTS est divisé en quatre jeux de données, chacun donnant de nouvelles informations sur le déplacement ou sur l’usager qui a effectué le déplacement. De nombreux facteurs sont également communs à plusieurs voire l’ensemble des jeux de données. Voici en résumé les jeux de données à disposition :

Tableau 2 : Ensemble des jeux de données disponibles sur le site du NHTS

Document	Facteurs observés	Description du contenu
Persons_pub	128	Ensembles de facteurs, principalement socio-économiques décrivant l’usager (<i>Persons</i>), mais décorrélés du déplacement et incluant des données liées au ménage
Households_pub	35	Ensembles de facteurs, principalement socio-économiques relatifs au ménage de chaque participant (<i>Households</i>) mais décorrélés du déplacement
Vehicules_pub	46	Ensembles de facteurs, principalement socio-économiques décrivant le(s) véhicule(s) au sein des ménages considérés, (incluant de fait les données liées à ces ménages) mais décorrélés du déplacement
Trip_pub	84	Ensembles de facteurs décrivant les différents déplacements (<i>Trips</i>) effectués par les l’usager interrogés, incluant les données liées aux ménages et à ces mêmes usagers (facteurs socio-économiques)

Ces jeux de données sont directement accessibles sur le site du *NHTS* ([National Household Travel Survey, 2024](#)) sous forme de fichiers texte qu’il faut ensuite mettre en forme sur Excel (voir par après). En plus des données brutes, l’étude fournit des documents explicitant l’ensemble des variables ou dénominations utilisées dans les jeux de données, avec un fichier général (*Dictionary*) et un fichier contenant l’ensemble des variables par jeux de données (*Codebooks*) sous la forme de fichiers Excel

(voir Figure 1 ci-après). A cela s'ajoutent un ensemble de notes techniques annexes expliquant le fonctionnement de l'étude et la collecte des données notamment.

3.1.2 Collecte des données

Ce travail de recherche vise uniquement à traiter et analyser les données recueillies par un organisme tiers, ici le NHTS. Il convient cependant d'explicitier la méthodologie de collecte de données de cet organisme pour pouvoir par la suite fixer un certaines nombres d'hypothèses ou de limitations lors de l'analyse. Le NHTS explicite l'ensemble de sa démarche au travers de notes explicatives et techniques relatives à la conduite de l'étude et la collecte des données. Ces documents sont également disponibles sur le site de l'organisme.

L'ensemble des données collectées par le NHTS l'a été par une combinaison de sondages téléphoniques et en ligne. Avant 2017, l'ensemble des données étaient recueillies sur papier ou bien par entrevue téléphonique. Depuis, l'organisme propose aux répondants de saisir leurs déplacements à l'aide d'un outil de cartographie, via une interface de programmation (API) basée sur une application compatible avec Google. Cette méthode de collecte constitue désormais le principal mode de collecte des données avec le développement d'internet et des nouvelles technologies. Une version courrier a été offerte pour ceux qui n'ont pas accès à Internet ou qui ne sont pas à l'aise avec cette technologie.

La collecte de données a été grandement simplifiée depuis 2017 afin de réduire le fardeau des répondants sans sacrifier les besoins de base en matière de données sur les comportements et les modes de déplacement. Les données fournies sont donc le fruit d'un ensemble de données complémentaires qui ont été croisées :

- les données de base traditionnelles recueillies au moyen d'une enquête sur les déplacements des ménages
- Les produits de données sur l'origine et la destination des passagers et des camions élaborés à l'aide de données passives provenant de sources du secteur privé

Les données de base donnent des informations sur les habitudes de déplacement qui documentent les tendances actuelles en matière de mobilité à l'échelle nationale, des divisions de recensement et des zones géographiques urbaines et rurales. La capacité d'établir des tendances en matière de paramètres de déplacement clés, comme les déplacements de personnes et les kilomètres parcourus, les déplacements en véhicule et les kilomètres parcourus en véhicule, l'utilisation du mode de transport, l'heure de la journée.

Il faut noter que cette étude de 2022 a recueilli moins d'enquêtes que les années passées (ex : 7 500 ménages à l'échelle nationale au lieu de 26 000 en 2017) ce qui constitue déjà une limitation car plus le jeu de données est faible et moins le modèle sera pertinent. C'est la raison pour laquelle l'ABS cherchera par la suite à croiser ces données traditionnelles avec un échantillon parallèle de 7 500 ménages basé sur modèles probabilistes pour en évaluer la pertinence. Cependant, seules les données traditionnelles seront considérées dans ce travail de recherche.

Les enquêtes ont été réparties sur les 7 jours de la semaine (jour précisé dans l'étude) sur l'ensemble des 365 jours de l'année (dates précises non précisées dans l'étude hormis le mois et l'année). Les questions de base, incluses ici, permettent de collecter des données démographiques et le comportement le jour du voyage (ex : le mode de transport, le but, l'heure de début et de fin du voyage, etc).

Le programme en ligne guide les répondants dans la déclaration des caractéristiques du ménage, puis, après avoir créé une liste des membres du ménage et des détails sur le véhicule, recueille les détails d'une journée de voyage auprès de chaque membre du ménage âgé de 5 ans et plus. Le programme, qui comprend une API Google (ou interface) pour cartographier les destinations de voyage, est programmé pour guider efficacement les répondants à travers les questions avec des *sauts intégrés* où les questions peuvent ne pas être applicables en fonction des réponses aux questions précédentes.

Les participants à l'étude reçoivent une incitation pour participer à l'étude. Pour l'ABS, un billet de 2 dollars est apposé sur la lettre d'invitation. Les ménages se voient offrir 10 dollars pour remplir le profil du ménage, puis 5 dollars pour chaque membre du ménage âgé de 5 ans et plus qui déclare ses déplacements pour la période de 24 heures.

Le NHTS 2022 utilise une approche d'échantillonnage à double base d'études parallèles utilisant des bases d'échantillonnage ABS et PFS. Dans le cas de l'ABS, base très ancienne, il s'agit d'échantillonner au hasard les ménages d'une région géographique donnée. La PFS, plus récente, utilise elle un échantillon choisi au hasard à partir d'un échantillon de panel en ligne établi basé sur des probabilités. Le NHTS 2022 a cherché à combiner ces deux bases complémentaires pour obtenir des données plus solides et complètes :

Tableau 3 : Forces et faiblesses des techniques d'échantillonnage opérées par le NHTS dans le cadre de l'étude

Bases	Avantages	Inconvénients
ABS	<ul style="list-style-type: none"> - zones géographiques ou des sous-populations difficiles d'accès - sont très efficaces compte tenu de leur nature non groupée - capte le gros des ménages (99%) 	<ul style="list-style-type: none"> - sous-représentation : personnes à faible niveau socio-économique, personnes dont l'anglais n'est pas la langue maternelle, minorités - l'augmentation du coût pour assurer un taux de réponse décent
PFS	<ul style="list-style-type: none"> - offre l'avantage de sauter la phase de recrutement initiale des ABS évitant la notion de consentement couramment observée dans les sondages en ligne. - qualité des échantillons de panel peut être similaire à celle des échantillons postaux bien conçus - taux de répondants plus élevés, un biais de non-réponse plus faible et des coûts réduits 	<ul style="list-style-type: none"> - recherches supplémentaires pour mesurer le comportement de voyage pour mesurer le comportement de voyage

L'utilisation d'application pour smartphone permet également d'atténuer l'incompréhension des répondants quant à la définition d'un voyage. Les répondants confondent souvent deux voyages ou plus avec un seul voyage et l'enregistrent comme tel. Cependant, les membres plus âgés ou plus jeunes du ménage qui n'ont peut-être pas leur propre smartphone seront exclus de l'échantillon ou invités à déclarer leurs déplacements à l'aide d'un sondage en ligne et cela peut limiter de fait le nombre de ménages pour lesquels les données peuvent être collectées sur plusieurs jours pour tous les membres du ménage.

Les chercheurs et les praticiens s'attendent à ce que le NHTS 2022 fournisse des informations précieuses sur les habitudes de voyage des ménages en cette période post-pandémique. Les questions de l'étude permettent à la fois de collecter des données traditionnelles sur les spécificités de chaque

voyage effectué, ainsi que des informations sur l'impact de la pandémie de COVID-19 sur les déplacements quotidiens, les voyages récents sur de longues distances ou encore l'utilisation de nouveaux modes de déplacement (covoiturage, trottinette électrique et vélos en libre-service).

En ce qui concerne les réponses au questionnaire, celles-ci sont encodées sous forme d'un entier allant de 0 à l'infini qui constitue un code de différenciation pour l'ensemble des réponses pour une variable donnée. En voici un exemple :

				04=Occupied without payment	275
HOUSEID	Unique Identifier- Household	C	10	Responses=9000013002-9000218040	31 074
LIF_CYC	Life Cycle classification for the household	C	2	01=one adult, no children	2 575
				02=2+ adults, no children	6 199
				03=one adult, youngest child 0-5	223
				04=2+ adults, youngest child 0-5	3 539
				05=one adult, youngest child 6-15	837
				06=2+ adults, youngest child 6-15	6 106
				07=one adult, youngest child 16-21	336
				08=2+ adults, youngest child 16-21	1 546
				09=one adult, retired, no children	1 844
				10=2+ adults, retired, no children	7 869
LOOP_TRIP	Trip origin and destination at identical location	C	2	01=Loop trip	351
				02=Not a loop trip	30 723
MSACAT	MSA category for the HH home address	C	2	01=MSA of 1 million or more, with rail	6 729
				02=MSA of 1 million or more, without rail	9 994
				03=MSA less than 1 million	12 896
				04=Not in MSA	1 455
MSASIZE	Population size category of the MSA from the five-year ACS API	C	2	01=In an MSA of Less than 250,000	5 529
				02=In an MSA of 250,000 - 499,999	3 078
				03=In an MSA of 500,000 - 999,999	4 289
				04=In an MSA or CMSA of 1,000,000 - 2,999,999	7 177
				05=In an MSA or CMSA of 3 million or more	9 546
				06=Not in MSA or CMSA	1 455
NONHHCNT	Number of non-household	N	8	Responses=0-98	31 074

Figure 1 : Extrait du Codebook avec l'ensemble des abréviations et le code associé à chaque variable étudiée (partie TRIP)

Certaines variables indiquent directement une quantité (ex : le nombre de personnes dans le foyer ou l'ID de la personne) et non une option associée à un entier. Cette quantité peut d'ailleurs être un nombre décimal, comme le nombre de miles parcourus. L'idée est d'avoir un fichier final constitué uniquement de valeurs numériques et non du texte afin de faciliter le traitement des données. Il faut également noter que les répondants peuvent choisir de ne pas répondre à l'une des questions qui leur sont posées et les valeurs suivantes sont inscrites à la place selon le contexte :

Tableau 4 : Réponses négatives de l'enquête (NHTS)

Valeur	Libellé	Description
-1	Saut valide	Quand la question n'est pas considérée de façon Données d'entrée appropriée par la personne interrogée et que par défaut aucune valeur n'existe
-7	Je ne préfère pas répondre	Information donnée par le répondant comme une source valide de donnée
-8	Je ne sais pas	Information donnée par le répondant comme une source valide de donnée
-9	Pas certain	Données d'entrée quand la question n'est pas considérée de façon inappropriée par la personne interrogée et que par défaut aucune valeur n'existe

Cela évite d'avoir des valeurs nulles dans le fichier final, ce qui facilite également le traitement des données et donne en plus des précisions sur le motif du refus.

3.2 Traitement des données

3.2.1 Mise en forme et filtrage des données de base

La mise en forme des données implique de connaître la méthode de traitement et surtout les outils utilisés pour l'analyse des données. Les études préliminaires ont permis de mettre en évidence des liens entre les variables étudiées ici et la littérature, voici les facteurs que l'on retrouve dans les données du NHTS :

Tableau 5 : Lien entre les données du NHTS et la littérature

Paramètre	Description	Lien avec la littérature
ADULTS	Nombre d'adultes (par ménage)	Lien direct (<i>facteur socio-économique</i>)
DRIVER	Conducteur ou non	Lien direct (<i>facteur socio-économique</i>)
DRIVERS	Nombre de conducteurs (par ménage)	Lien direct (<i>facteur socio-économique</i>)
EDUC	Niveau d'éducation	Lien direct (<i>facteur socio-économique</i>)
EMPLOYMENT	Statut salarial	Lien direct (<i>facteur socio-économique</i>)
HHFAMINC	Revenus du foyer	Revenu à l'échelle du foyer et non individuelle mais lien quand même (<i>facteur socio-économique</i>)
HOUSEVEH	Nombre de personnes (par ménage)	Lien direct (<i>facteur socio-économique</i>)
LIF_CYC	Composition de la famille	En lien avec le voyage avec d'autres personnes (<i>facteur lié au contexte</i>)
PRMACT	Situation professionnelle	Lien direct (<i>facteur socio-économique</i>)
R_Age	Age des usagers	Lien direct (<i>facteur socio-économique</i>)
R_HISP	Origine hispanique	Lien direct (<i>facteur socio-économique</i>)
R_RACE	Origine ethnique	Lien direct (<i>facteur socio-économique</i>)
R_SEX	Sexe de la personne	Lien direct (<i>facteur lié au contexte</i>)
TDAYDATE	Mois et année du trajet	Lien important en fonction de la période de l'année (<i>facteur lié au contexte</i>)
TRAVDAY	Jour de la semaine	Lien direct (<i>facteur lié au contexte</i>)
TRPMILES	Distance parcouru (continue)	Lien direct (<i>facteur liés aux transports</i>)
TRVCMIN	Temps de trajet (continue)	Lien direct (<i>facteur liés aux transports</i>)
URBANSIZE	Densité de population	Lien direct (<i>facteur lié au contexte</i>)
URBRUR	Rural/urbain	Lien direct (<i>facteur lié au contexte</i>)
WHYTO	But du voyage	Lien direct (<i>facteur lié au contexte</i>)

Ces données seront utilisées lors de la phase d'analyse pour déterminer les facteurs ayant une influence sur la multimodalité des usagers des transports aux Etats-Unis. A cela s'ajoutent d'autres variables collectées par le NHST au sujet des déplacements qui seront également utilisées par la suite pour créer de nouvelles variables d'intérêt.

Les données ont subi un premier traitement sous Excel afin notamment de croiser les différents fichiers de données fournis par le NHTS et de pouvoir également mieux visualiser ces données dans un même tableur. Le logiciel d'analyse statistique R a par la suite été privilégié pour l'essentiel du traitement des données et d'analyse car il est parfaitement adapté à l'analyse statistique des données. Une partie de

la mise en forme des données a également été réalisée sur ce logiciel en amont de l'analyse. Les données de base auraient pu être directement analysées sur R dès le départ mais le passage par Excel permet d'avoir un fichier de base plus lisible qui facilite grandement le travail par après sur le logiciel R.

Comme dit précédemment, les fichiers texte (ou CSV) du NHTS ont subi une première conversion afin de séparer les différentes catégories des données de base en différentes colonnes séparées. Pour cela, l'identificateur de texte guillemets et la virgule ont été utilisés afin de séparer les différentes catégories ou paramètres de l'étude. Des quatre fichiers de base, le fichier relatif aux déplacements *Trip_pub* est le plus complet car, d'une part, il combine des données de différents types (facteurs sociaux économiques, propre au véhicules utilisés ou encore au trajet en lui-même). D'autre part, il affiche quatre identifiant que l'on retrouve dans les autres fichiers qui va permettre de faire la correspondance entre les table :

- L'ID de l'utilisateur (ou *PERSONID*)
- L'ID du déplacement (ou *TRIPID*)
- L'ID du foyer (ou *HOUSEID*)
- L'ID du véhicule éventuel utilisé lors du trajet (ou *VEHCASEID*)

Devenu le fichier de référence, *Trip_pub* est complété des catégories en provenance des trois autres fichiers (*Persons_pub*, *Households_pub* et *Vehicles_pub* tant que celles-ci ne sont pertinentes avec l'étude à mener et non redondantes avec les catégories existantes. Voici les données supprimées dans le fichier de trajets (*Trip_pub*) :

Tableau 6 : Données de *Trip_pub* supprimées

Données non conservées		
Paramètre	Description	Justification
FRSTHM	A commencé la journée de voyage à la maison	Paramètre non étudié
PARK	Stationnement payant à tout moment pendant la journée de voyage	Paramètre non étudié
TRIPPURP	Objectif général du voyage	Redondance avec <i>WHYTO</i>
WHYTRP1S	Récapitulatif de l'objet du voyage	Redondance avec <i>WHYTO</i>
WHODROVE_IMP	Personne présumée qui a conduit pendant le voyage	Redondance avec <i>WHODROVE</i>
PARK2_PAMOUNT	Montant payé pour le stationnement	Paramètre non étudié
PARK2_PAYTYPE	Périodicité du paiement du stationnement	Paramètre non étudié
PARK2	A payé pour le parking	Paramètre non étudié
WTTRDFIN	Poids du voyage sur 7 jours	Chaque usager est regardé indépendamment
WTTRDFIN5D	Poids du voyage sur 5 jours	Chaque usager est regardé indépendamment
WTTRDFIN2D	Poids du voyage sur 2 jours	Chaque usager est regardé indépendamment
VMT_MILE	Distance de trajet calculée (miles) pour les trajets du conducteur	Redondance avec <i>TRPMILES</i>

GASPRICE	Prix hebdomadaire régional de l'essence, en cents, au cours de la semaine de la le jour de voyage du ménage ;	Paramètre non étudié
WHYTRP90	Objectif de l'excursion d'une journée conforme à la conception du SNPT de 1990	Paramètre non étudié
HOMEOWN	Qu'il s'agisse d'une maison ou d'une location	Paramètre non étudié
RAIL	Statut de train lourd MSA pour les ménages	Paramètre non étudié
CENSUS_D	Classification <i>D</i> de l'adresse du domicile	Paramètre non étudié
CENSUS_R	Classification <i>R</i> de l'adresse du domicile	Paramètre non étudié
CDIVMSAR	Regroupement des ménages selon la combinaison de la division de recensement, du statut MSA, et la présence de voies ferrées	Paramètre non étudié
MSACAT	Catégories MSA pour l'adresse du ménage	Paramètre non étudié
PROXY	Sondage rempli par soi-même ou par quelqu'un d'autre	Paramètre non étudié
R_SEX_IMP	Sexe du répondant (imputé)	Redondance avec <i>R_SEX</i>

Les données supprimées dans les autres fichiers sont indiquées dans les annexes (Tableau 39, Tableau 40 et Tableau 41). Le fichier obtenu après la fusion des quatre fichiers originels est utilisable en la matière sous R mais très peu visible pour un utilisateur lambda cherchant à le lire. En effet, il apparaît comme un tableau constitué uniquement de valeurs numériques non discernables sans la table de correspondance associée (ou *Codebooks*). Certaines colonnes vont donc être dédoublées pour afficher la valeur de texte correspondante à chaque valeur numérique grâce notamment aux fonctions *RechercheV* et *Si* propre à Excel. On retrouve alors les correspondance suivantes :

Tableau 7 : Correspondances entre les paramètres numériques et leur équivalent sous format texte

Paramètres de base	Description	Paramètres créés
TDAYDATE	Date du jour du voyage (AAAAMM)	DATE
TRPTRANS	Modes de transport dérivés	TR_TYPE
DRIVER	Distance de trajet calculée (miles) pour les trajets du conducteur	DRIVER_STATUS
LIF_CYC	Composition familial des ménages	FAMILY
R_SEX	Sexe du répondant	SEXE
EDUC	Niveau d'éducation le plus élevé	EDUCATION
HHFAMINC	Revenus du ménage	HH_INCOME
R_RACE	Origine ethnique du répondant	RACE
R_HISP	Origine Hispanic ou non du répondant	HISP
WORKER	Situation professionnelle du répondant	EMPLOYMENT
URBAN	Classification des zones urbaines des ménages, basée sur le fichier de formes TIGER/Line 2020	URBN
URBANSIZE	Zone urbaine ou le domicile est localisé	DENSITY
URBRUR	Ménage en zone urbaine/rurale	URBANIZED

WHYTO	Raison du voyage jusqu'à la destination	PURPOSE
TDWKND	Trajet le weekend	TIME_WEEK
PRMACT	Activité principale pour ceux qui n'ont pas travaillé pour un salaire la semaine dernière	ACTIVITY

Toujours dans une optique de mise en forme, les temps de début et de fin de trajet ont été mis au format heure. Ces ajustements restent par ailleurs réalisables directement via le logiciel R.

3.2.2 Prétraitement des données et définition de la multimodalité

Microsoft EXCEL

L'essentiel des modifications apportées aux données ont été réalisées à l'aide du logiciel R. En revanche, un identifiant unique pour chaque usager (*PER_ID*) a été créé directement sur Excel en prenant en compte une partie de l'ID du déplacement (ou *TDCASEID*) qui est commune à l'ensemble des trajets effectués par l'usager considéré. Seuls les deux derniers digits du déplacement changent pour indiquer que l'on passe d'un trajet à un autre dans le tableau (ex : *TDCASEID = 13042010*, et *PER_ID= 1304201*). Cette opération sera à la base de la création d'un tableau des usagers par la suite sur R.

A cela s'ajoute la suppression des quatre premiers digits de l'ID des déplacements (et donc par extension et des personnes) dans le but de simplifier visuellement le jeu de données. Ainsi, la séquence *9000* commune à tous les trajets est supprimée (ex : passage de *9000130020101* à *130020101*). Enfin, une colonne vide *MULTIMODAL* a été créée avant le passage sur R pour anticiper l'indication d'un éventuel trajet multimodal (*Yes/Oui* si multimodal). Le fichier est alors prêt à être importé sur R.

R studio

Le fichier Excel des déplacements peut directement être importé sur R où il sera désormais considéré comme le nouveau fichier de référence (*data_TRIP*). L'ensemble des valeurs négatives du tableau (soit toutes les réponses négatives des participants) ont été supprimées afin de faciliter le traitement des données par la suite. La différenciation de plusieurs types de réponses négatives par le NHTS (voir Tableau 4) est une variable intéressante dans l'absolue mais elle ne constitue pas pour autant un intérêt pour cette analyse.

Ce fichier, comptant 65 indicateurs, comptabilise toujours l'ensemble des déplacements effectués durant la journée par les personnes interrogées (31074 au total) en indiquant également la personne qui l'a effectué (*PERSONID*). Ainsi on retrouve une ou plusieurs lignes pour chaque usager en fonction du nombre de déplacements effectué durant la journée où il a été interrogé et les trajets sont notés à la suite dans le tableau (Figure 2 en annexe).

L'objectif de l'étude est de montrer si les usagers sont multimodaux et de déterminer les facteurs qui influencent la multimodalité de ces personnes. La séquence de déplacement est ici une succession d'au moins deux déplacements qui ne se suivent pas dans le temps (plus de 5 minutes d'écart selon les critères du NHTS) et ce même s'ils sont effectués par la même personne et avec le même mode de transport. Ainsi pour chaque utilisateur est associée une séquence de transport qui compile l'ensemble de ses déplacements à la journée :

Tableau 8 : Séquence de transports sur une journée (10 premières personnes interrogées)

PER_ID	Trip_Sequence	Vehicle_Sequence
1300201	1,2	SUV/Crossover, SUV/Crossover
1300202	1,2,3	SUV/Crossover, SUV/Crossover, SUV/Crossover
1300204	1,2	SUV/Crossover, SUV/Crossover
1301601	1,2,3	Car, Car, Car
1301602	1,2,3	Car, Car, Car
1303901	1,2	Car, Car
1304201	1,2	SUV/Crossover, SUV/Crossover
1304202	1,2	SUV/Crossover, SUV/Crossover
1304301	1	Walked
1304401	1,2	Walked, Walked

L'étude du NHTS précise que le trajet est un déplacement de type boucle (variable *Loop*) donc techniquement deux fois le même déplacement successivement. Or, méthode d'analyse explicitée par après n'a pas besoin de dupliquer ce type de trajet dans le tableau global des déplacements, il ne sera donc pas considéré.

Dans un premier temps, la multimodalité des personnes interrogées sera faite à l'échelle du trajet dans sa globalité, soit l'ensemble des déplacements effectués durant la journée (ou séquence des déplacements, noté *Vehicle_Sequence*). La définition du caractère multimodal d'un trajet nécessite de poser un certain nombre d'hypothèses cohérentes basées en partie sur la littérature et sur les limites fixées par l'étude.

Modes de transport et multimodalité :

- Ce travail vise à étudier la multimodalité dans les centres urbains et sur des distance relativement courtes et de fait, les déplacements effectués en avion ont un impact **négatif** sur la multimodalité des usagers
- Les déplacements effectués par le biais un véhicule personnel motorisé à quatre roues (ex : voiture, SUV, Van, etc.) ont un impact **négatif** sur la multimodalité des usagers
- Les autres modes de transport (*Others*) ne sont pas explicités, mais cela inclut sans doute les modes exotiques de proximité (ex : les monoroues, les skateboards ou encore les trottinettes) et ces déplacements ont donc un impact **positif** sur la multimodalité des usagers
- Les déplacements effectués par le biais des transports en commun (ex : bus, train, etc) des ont un impact **positif** sur la multimodalité des usagers
- Les déplacements effectués par le reste des modes de transport (ex : deux-roues, taxi, voiture en libre-service, etc) ont un impact **positif** sur la multimodalité des usagers
- Le cas de la marche est ambigu car une catégories spécifique du tableau de données (*Walk*) indique si l'utilisateur a marché entre sa voiture (le parking) et un autre mode de transport (ex : les transports en communs). De plus, la marche est toujours présente lors du passage d'un mode de transport à un autre, elle ne sera pas considérée dans l'analyse par après

Tableau 9: Tableau récapitulatif des catégories de modes de transport et leur impact sur la multimodalité

Type de véhicule	Impact
Car (voiture) [1]	Négatif
SUV/Crossover [1]	Négatif
Pickup truck [1]	Négatif
Recreational Vehicle (véhicule de loisir) [1]	Négatif
Van [1]	Négatif
Amtrak (tramway) [2]	Positif
Commuter rail (train de transit) [2]	Positif
Public or commuter bus (bus public ou de transit) [2]	Positif
School bus (bus scolaire) [2]	Positif
Streetcar or trolley car (système de navette) [2]	Positif
Subway or elevated rail (métro ou train aérien) [2]	Positif
Airplane (avion) [3]	Non considéré
Walked (marche) [3]	Non considéré
Bicycle (including bikeshare, ebike...) (tous types de vélos) [4]	Positif
E-scooter (scooter électrique) [4]	Positif
Motorcycle (moto, scooter) [4]	Positif
Other ride-sharing service (système de partage, covoiturage) [4]	Positif
Other (specify) (autre) [4]	Positif
Paratransit/ Dial a ride (service de véhicule pour les PMR) [4]	Positif
Taxicab or limo service (taxi ou service de limo) [4]	Positif

Avec : les véhicules individuels quatre roues ([1]), transports en commun ([2]), modes exclus ([3]), autres modes ([4]).

Il faut également poser d'autres hypothèses relatives à la séquence même des déplacements et ce pour chaque usager. Ainsi un trajet est considéré comme étant multimodal à l'échelle de la journée si et seulement si l'ensemble des conditions suivantes sont réunies :

- La séquence contient au moins deux éléments uniques (ex : voiture, voiture, vélo, voiture)
- La séquence contient au moins un élément ayant un impact positif sur la multimodalité (voir tableau précédent)
- Les déplacements en boucle (*Loop*) ne sont pas comptés deux fois comme expliqué précédemment

Cette première approche permet déjà d'avoir une estimation du nombre de personnes multimodales à l'échelle de la journée, mais cette méthode possède des limitations. En effet, certaines personnes peuvent être considérées comme multimodales avec des déplacements très espacés temporellement les uns des autres (ex : une personne se rendant au travail en train le matin et qui revient en taxi le soir chez elle). La marche reste problématique car elle est considérée de façon ambiguë dans l'étude comme mode spécifique (pour se balader, se déplacer) alors qu'elle est censée être présente à chaque étape de la séquence (ex : pour aller à l'arrêt de bus ou au parking pour récupérer sa voiture). Dans tous les cas, le fichier des usagers (*data_TRIP_Persons*) est complété par un nouveau paramètre *Multimodalité* (*Multimodality*) qui prend les valeurs binaires Oui (*Yes*) ou non (*No*) pour indiquer si la séquence est multimodale et par extension l'utilisateur.

Une autre approche consiste à vérifier, pour chaque usager, la présence d'une présence éventuelle d'une séquence multimodale sur deux déplacements consécutifs plutôt que sur le trajet total (somme des déplacements). Les critères pour définir une séquence multimodale seraient les mêmes que dans le cas du trajet sur la journée et on s'attend à retrouver une proportion de personnes multimodales

équivalente dans les deux cas. Ce travail sera effectué sur le fichier *data_TRIP_Final* qui correspond donc à l'équivalent du fichier de base des déplacements mais en prenant en compte toutes les modifications apportées au fichier des personnes (*data_TRIP_Persons*).

3.2.3 Ajustement des variables observées

La multimodalité constitue le paramètre central de ce sujet de recherche mais les variables identifiées dans le Tableau 5 : Lien entre les données du NHTS et la littérature doivent être ajustées afin de faciliter l'analyse statistique qui suit. Chaque variable est associée à un certain nombre de réponses possibles et certaines réponses se retrouvent beaucoup plus fréquemment que d'autres (ex : un nombre négligeable de personnes Hawaïenne en comparaison de du nombre de personnes blanches). Il convient donc de réduire le nombre de catégories de réponses possibles en regroupant les catégories contenant le moins de réponses tout en conservant une certaine cohérence et en restant en accord avec la littérature (cf. partie *ajustement des variables* en annexe).

L'origine ethnique des répondants a été séparée en deux variables : la race (*R_RACE*) et l'origine hispanique ou non du répondant (*R_HISP*). La population d'origine hispanique constitue la deuxième population la plus représentée aux Etats-Unis après la population blanche majoritaire. Celle-ci est même majoritaire dans certaines villes ou états du Sud du pays (Nouveau Mexique, Floride), ce qui peut expliquer en partie son statut à part. Cela s'explique également par le fait que certaines personnes d'origine hispanique se définissent comme blanc de peau (*White*) ou comme ayant une double origine (*Multiple races selected*).

La variable *Hispanique* a finalement été incorporée comme simple réponse au sein de la catégorie *Race* et remplace la valeur qui existait alors (principalement *White*). Du fait de leur très faible nombre, les amérindiens (*American Indian/Alaska Native* et *Native Hawaiian/Pacific Islander*) ont été mélangés avec les personnes avec plusieurs origines (*Multiple race selected*) au sein de la catégorie existante *Autre race (Other race)*.

Pour la variable Sexe (*R_SEX*), les deux réponses principales *hommes (Males)* et *femmes (Female)* ont été conservées. Les réponses des personnes ne souhaitant pas répondre (*don't know, Refuse*) ont été réunies dans la catégorie *Other sex*, cas marginal mais à considérer.

La possession d'un emploi (*EMPLOYMENT*) a été concaténée avec le type d'activité (*ACTIVITY*) comme nouveau type de réponse (*Worker*). Les personnes en recherche d'emploi (*Looking for work/unemployed*) ou temporairement au chômage (*Temporarily absent from a job or business*) sont considérées comme sans emploi (*Unemployed*). Enfin, la catégorie *Something else* a été renommée *Autre activité (Other Activity)* sur le modèle des autres variables étudiées.

Concernant l'éducation (*EDUC*), les réponses liées au lycée ont été réunies (*Some high school, no diploma/GED, High school graduate, have diploma/GED*), de même pour celles liées au Collège américains (*Some college, no degree/ some trade school, Associates degree (2-year)/ trade school certificate*).

La densité urbaine (*URBANSIZE*) a été réduite à trois grandes catégories de distance (*50,000-499,000* habitants, *500,000-999,999* habitants et *1,000,000 or more* habitants) et une quatrième réponse *hors zone urbanisée (Not in urbanized area)* est renommée *Other density*. Cela permet notamment de distinguer rapidement les personnes vivant dans les grandes villes de celles vivant dans les pôles urbains secondaires voire à la campagne (zones très peu denses).

Sur le même modèle, la variable *Revenus du foyer* (*HH_INCOME*) a été réduite à trois grandes catégories (*Less than \$24,999*, *\$25,000 to \$100,000* miles et *\$100,000 or more*).

Dans le cas de la composition familiale (*LIF_CYC*), le fait d'être à la retraite n'est pas jugé comme pertinent et les catégories *Sans enfants* (*no children*) seront donc réunies selon le nombre de parents dans le foyer (un ou deux). L'âge du permis de conduire (16 ans aux Etats-Unis) constitue une démarcation claire pour différencier un enfant d'un adulte. Les catégories d'âge allant de 0 à 15 ans ont donc été réunies que cela soit pour un ou deux adultes.

Pour ce qui est du but du déplacement (*WHYTO*), les catégories relatives au travail ont été réunies dans une catégories unique *Work* même s'il s'agit parfois de bénévolat (*Work from home*, *Work at a non-home location*, *Work activity to drop-off/pickup someone/something*, *Volunteer activities (not paid)*). Les réponses relatives à la santé ont été réunies dans la catégorie *Health* (*Attend childcare or adult care* et *Health care visit*). Celles relatives aux achats sont concaténées dans la catégorie *Spending*s (*Buy meals*, *Shop/buy/pick-up or return goods*, *Other family/personal errands*). Enfin une dernière catégorie *Recreational*) réunit les réponse des activités liées aux loisirs (*Exercise*, *Rest or relaxation/vacation*, *Religious or other community activities*)

Le nombre de conducteurs dans le foyer (*DRVRCNT*) est réduit à 5 réponses et dont la dernière catégorie concerne tous les foyers avec plus de 4 conducteurs (*4+* conducteurs). Dans la même idée, le nombre d'adultes (*NUMADLT*) est limité à 5 réponses avec une réponse *4+* si le foyer compte au moins 4 adultes ou plus. La logique est toujours la même pour le nombre de personnes dans le foyer (*HHVEHCNT*).

Aux variables existantes s'ajoutent des variables créées de toute pièce pour servir d'alternative à certaines variables quasi continues (ex : l'âge des participants). Ces variables discrètes se présentent sous la forme d'intervalles construits en grande partie grâce à la littérature : *Travel_time_Intervals*, *Distance_Intervals* et *Age_Intervals*.

Travel_time_Intervals, ou intervalles pour les temps de trajet : cette variable prend 11 valeurs d'intervalles différentes entre les valeurs limites 0 min et 1140 min (le maximum de l'étude). Le temps de transport est indiqué pour chaque déplacement (*TRVLCMIN*), il suffit alors de faire la somme de ses temps de transport pour chaque participant afin d'obtenir le temps de trajet total sur une journée. Ce durée totale est ensuite comparée aux différents intervalles de temps créés : *(0,5]*, *(5,10]*, *(10,15]*, *(15,20]*, *(20,30]*, *(30,45]*, *(45,60]*, *(60,120]*, *(120,240]*, *(240,360]*, *(360,1140]*. Le nombre important d'intervalles pour les faibles durées s'explique par le fait que les trajets en ville sont plutôt courts et on s'attend à ce que la multimodalité soit plus importante.

Distance_Intervals, ou intervalles pour la distance parcourue lors du trajet : prend 8 valeurs d'intervalles différentes entre les valeurs limites 0 miles et 5000 miles (le maximum de l'étude). La distance parcourue est également indiquée pour chaque déplacement (*TRPMILES*), il suffit alors de faire la somme pour chaque participant ce qui permet d'obtenir la distance totale parcourue sur une journée qui est comparée aux différents intervalles créés : *(0,0.5]*, *(0.5,2]*, *(2,5]*, *(5,10]*, *(10,25]*, *(25,50]*, *(50,100]*, *(100,5000]*. Le nombre important d'intervalles pour les faibles distances s'explique par le fait que les trajets en ville sont plutôt courts, notamment ceux effectués à pied et on s'attend à ce que la multimodalité soit plus importante pour ces trajets.

Age_Intervals prend 7 valeurs d'intervalles différentes entre les valeurs limites 5 an et 92 ans (le maximum de l'étude). L'âge des participants (*R_AGE*) est alors comparé aux différents intervalles créés : *(5,13]*, *(13,16]*, *(16,25]*, *(25,35]*, *(35,50]*, *(50,65]*, *(65,92]*. La limite de 16 ans est importante car elle constitue l'âge minimal légal pour conduire aux Etats-Unis. Les autres intervals ont été choisis pour

distinguer les jeunes adultes (<25 ans) encore en étude de ceux qui sont au début de leur carrière. La limite de 65 ans correspond à celui de la retraite officielle aux Etats-Unis.

Tableau 10 : Récapitulatif des changements

Variables d'entrée	Nombre de sous catégories en entrée	Variables de sortie	Nombre de sous catégories en sortie
DRVRCNT	7	DRVRCNT	5
EDUC	8	EDUC	7
HH_INCOME	13	HH_INCOME	4
HHVEHCNT	11	HHVEHCNT	5
LIF_CYC	10	LIF_CYC	7
NUMADLT	8	NUMADLT	5
PRMACT+ EMPLOYMENT	7	PRMACT	6
R_RACE + HISPANIC	8	R_RACE	5
R_SEX	4	R_SEX	3
URBANSIZE	6	URBANSIZE	4
WHYTO	20	WHYTO	10

3.2.4 Préparation spécifique en vue de la régression logistique

Les changements précédents ont été effectués pour limiter le nombre de facteurs considérés dans l'analyse mais en accord avec la littérature tout en conservant une certaine variété dans les réponses. L'analyse des données via une régression logistique semble pertinente pour montrer l'influence de différents facteurs sur la variable *Multimodalité*. Il nous faut concaténer davantage les catégories propre à chaque variable observée afin d'améliorer les résultats du modèle et ainsi faciliter l'interprétation qui en découle. On veillera cependant à conserver un minimum de sens et un lien avec les informations fournies par la littérature.

La réduction du nombre de catégories a aussi été rendue possible par l'exécution d'une régression logistique générale entre les variables considérées (ex : l'âge) et la variable d'intérêt principal (*Multimodalité*). La régression aura donc la forme suivante (extrait du code R) :

```
Logistic_reg_Tot <- glm(Multi ~ R_SEX + WHYTO + DRIVER + TRAVDAY + Travel_time_Intervals + Distance_Intervals + URBANSIZE + HOUSEVEH + R_AGE + TRPMILES, data = data_TRIP_Person_V2, family = binomial(link = "logit"))
```

Seules les sous-catégories significatives (dont la p-value est la plus faible, soit proche de 0) ont été conservées comme telles. Les autres sous-catégories ont alors été regroupées entre elles par la suite (cf. partie *Ajustement complémentaires propres à la régression logistique* en annexe).

La variable *Age_Intervals* créée à partir de l'âge des participants affichait plus d'intervalles pour les jeunes participants que pour les personnes âgées. Cela permettait de montrer les changements de comportements liés à l'acquisition d'une voiture (permis de conduire à partir de 16 ans) ou bien le changement d'établissement scolaire (de l'école élémentaire au collège à 13 ans). Dans un souci de simplification, on ne conservera uniquement les catégories suivantes : (5,16[, (16,35[, (35,50[, (50,92[.

Pour ce qui est des intervalles liés au temps de trajet (*Travel_time_Intervals*) ou à la distance (*Distance_Intervals*), les catégories les plus faibles ont été davantage regroupées car le nombre de répondants est plus faible que pour les grandes catégories. Pour le temps de trajet, il n'y a plus que

deux intervalles de 30 minutes pour des temps inférieur à 1 heure ($(0,15[$ et $(15,30[$), le reste est inchangé. Pour la distance, la première catégorie regroupe toutes les personnes ayant un trajet inférieur à 10 km ($(0,10[$).

La régression logistique simple effectuée sur l'activité (*WHYTO*) a révélé les sous catégories suivantes comme significatives : *Regular activities at home* (tâches quotidiennes), *Attend school as a student* (aller à l'école en tant qu'étudiant) et *Change type of transportation* (changer de mode de transports). Ces sous-catégories ont donc été conservées par après. Cependant, les catégories relatives au travail (*Work*), à la santé (*Health*) ou au fait de déposer quelqu'un (*Drop off/pick up someone (personal)*) ont été regroupées dans *Other Mandatory* (autres obligatoires) par opposition aux déplacements non obligatoire : *Spending* (dépenses diverses), *Recreational* (activités récréatives), *Visit friends or relatives* (visite de proches).

Tableau 11 : Récapitulatif des derniers changements pré-régression logistique

la variable d'entrée/sortie	Nombre de sous catégories en entrée	Nombre de sous catégories en sortie
Age_Intervals	7	4
Distance_Intervals	5	4
Travel_time_Intervals	6	4
WHYTO	10	5

On notera par ailleurs que la variable *TDATE* n'est pas considérée ici car l'étude de base n'indique par la date précise mais simplement le mois et l'année du déplacement alors qu'il faut reconstituer le trajet sur une journée. Une date plus précise aurait servi à comparer la multimodalité des répondants entre les périodes de travail et les vacances ou jours fériés notamment.

3.3 Variables retenues et tests statistiques préliminaires

Le travail précédemment a permis d'obtenir un nouveau facteur binaire *Multimodalité* indiquant si la personne est multimodale ou non. Ce travail de recherche vise à mettre en évidence les facteurs qui ont un impact sur la multimodalité des personnes interrogées. Seront considérés, les 20 indicateurs suivants pour la suite de l'analyse :

Tableau 12 : Indicateurs considérés dans l'analyse et association à la variable d'intérêt principale

Variables	Nombre de catégories finales	Description	p-value
Variables discrètes (PEARSON Test)			
ADULTS	5	Nombre d'adultes dans le foyer	0,06
Age_Intervals	4	Intervalles liés à l'âge	<0,01
Distance_Intervals	4	Intervalles liés à la distance parcourue	<0,01
DRIVER	2	Possession d'un permis de conduire	<0,01
DRIVERS	5	Nombre de conducteurs dans le foyer	<0,01
EDUC	7	Niveau d'éducation de la personne	<0,01
HHFAMINC	5	Revenus du foyer	<0,01
HOUSEVEH	5	Nombre de véhicules dans le foyer	<0,01
PRMACT	6	Activité de la personne	<0,01
LIF_CYC	7	Composition familiale	<0,01
R_RACE	5	Origine ethnique de la personne	0,04
R_SEX	3	Sexe de la personne	<0,01
TRAVDAY	7	Jour du trajet	<0,01
Travel_time_Intervals	4	Intervalles liés au temps de trajet	<0,01
URBANSIZE	4	Densité	<0,01
URBRUR	2	En zone rurale ou urbaine	0,48
WHYTO	5	But du voyage	<0,01
Variables quasi continues (WALD Test)			
R_AGE	87	Age de la personne	<0,01
TRPMILES	8500	Distance parcourue	<0,01
TRVCMIN	320	Temps de trajet	<0,01

Avec en gras les variables d'intérêt à conserver quoi qu'il arrive (potentially confounding variables).

Des tests préliminaires ont été réalisés en amont de la régression logistique pour montrer la présence d'un lien entre la variable principale (ici la multimodalité ou *Multi*) et la variable considérée. Un test de Wald a été appliqué aux variables considérées comme quasi continues (*R_AGE*, *TRPMILES*, et *TRVCMIN*) car elles comptent un très grand nombre de valeurs individuelles. Un test de Pearson (ou test du Chi-carré) est quant à lui appliqué à toutes les autres variables qui ont été limitées à une poignée de catégories. Il s'agit alors de tester la probabilité d'observer une statistique du khi-deux aussi extrême, voire plus, que celle observée sous l'hypothèse nulle : tous les coefficients de régression du modèle sont égaux à zéro.

Dans tous les cas, une *p-value* faible (qui tend vers 0) indique une forte signification statistique et qu'il existe des coefficients de régression significativement différents de zéro. En termes pratiques, une *p-value* < 0,05 signifie que les données observées sont très incohérentes avec l'hypothèse nulle (pas de lien entre les variables du test), que les variables explicatives incluses dans le modèle ont un effet significatif sur la variable dépendante, dans notre cas la multimodalité ([Proc Logistic | SAS Annotated Output, 2024](#)).

On peut dire que chaque variable considérée dans le modèle a un effet significatif sur la variable dépendante (ici la multimodalité) et qu'il existe donc une grande association entre les deux variables de chaque test. En revanche, on ne peut pas dire à quel point la variable considérée a une influence sur la variable dépendante. Dans notre cas, la plupart des variables présentent une *p-value* inférieur à

0,05 indiquant une forte corrélation entre ces variables et multimodalité. La variable *ADULT* (nombre d'adultes dans le foyer) est légèrement au-dessus de 0,05 mais la corrélation est quand même suffisante. Par contre, la corrélation est relativement faible pour *URBRUR* (personne rurale ou urbaine).

Le degré d'association peut ensuite être évalué par le biais du test de Cramer au travers d'un table de contingence :

Tableau 13 : Récapitulatif des derniers changements pré-régression logistique

Variables	p-value	Cramer's V
ADULTS	0,04	0,03
Age_Intervals	<0,01	0,17
Distance_Intervals	<0,01	0,04
DRIVER	<0,01	0,19
DRIVERS	<0,01	0,10
EDUC	<0,01	0,15
HHFAMINC	<0,01	0,05
HOUSEVEH	<0,01	0,13
PRMACT	<0,01	0,15
LIF_CYC	<0,01	0,08
R_AGE	<0,01	0,21
R_RACE	0,05	0,03
R_SEX	0,01	0,03
TRAVDAY	<0,01	0,08
Travel_time_Intervals	<0,01	0,09
TRPMILES	1	0,97
TRVCMIN	NA	NA
URBANSIZE	<0,01	0,04
URBRUR	0,44	0,01
WHYTO	<0,01	0,26

Avec en gras les variables d'intérêt à conserver quoi qu'il arrive. On notera que la variable liée au temps de trajet (*TRVCMIN*) comporte trop de valeurs par rapport aux autres catégories et a donc été écartée.

La p-value est calculée cette fois-ci pour un test de vraisemblance (*likelihood ratio*) et non un test de Pearson (R: [Cramer's V, 2024](#)). Il s'agit du test du chi carré du rapport de vraisemblance selon lequel au moins un des coefficients de régression des prédicteurs n'est pas égal à zéro dans le modèle. Finalement, on retrouve les mêmes conclusions que lors du précédent Pearson car toutes les variables semblent associées à la variable *Multimodalité* hormis *URBRUR*.

En revanche, on note que le degré d'association est modérément faible (valeur de Cramer < 0,3 pour la plupart) voire très faible (< 0,05) pour une bonne partie des variables. Cela est potentiellement dû à la très faible proportion de personnes multimodales dans la population totale (moins de 2,8%). Les résultats sont donc dilués dans une population qui est globalement non multimodale, ce qui rend difficile la mise en évidence d'une association entre toutes ces variables et la multimodalité des usagers. L'association est en revanche plus forte pour la distance (*TRPMILES*) car là on atteint presque 100% d'association.

Ces tests permettent d'avoir une idée sur le niveau d'association entre les variables retenues et la variable d'intérêt principal mais seule une analyse plus poussée permet véritablement de mettre en évidence une corrélation entre elles (voir partie d'après).

3.4 Régression logistique et corrélation entre les variables

3.4.1 Principes et fonctionnement du modèle

L'objectif de cette étude est d'établir une corrélation entre les 20 variables précédemment retenues et la variable d'intérêt principal (la multimodalité). On cherche plus spécifiquement à déterminer si un utilisateur donné est multimodal ou non selon un critère considéré (en lien avec les variables retenues). La régression logistique semble être un modèle très bien adapté pour ce genre de situation où la variable d'intérêt est binaire et où seuls deux résultats sont possibles (0 pour *non multimodal*, constituant un échec ; ou 1 pour *multimodal*, constituant un succès). Elle modélise de fait la probabilité qu'une observation donnée appartienne à une catégorie particulière et où la probabilité comprise entre 0 et 1 ([Logistic Regression Essentials in R - Articles - STHDA, 2018](#) ; [Qu'est-ce que la régression logistique ?, 2024](#)).

Ce genre de modèle permet de trouver les relations entre deux variables données afin de prédire par la suite la valeur de l'un de ces facteurs en fonction de l'autre. La régression logistique est généralement moins complexe que les autres méthodes d'analyse de données (ex : arbres de décision, réseaux neuronaux, etc.) et est donc plus rapide à utiliser même sur des grands volumes de données (ici, 20 critères appliqués à 10 591 participants dans l'étude) car elle nécessite moins de capacité de calculs. Cela est d'autant plus pertinent que la performance du logiciel R ne permet pas l'utilisation de modèles de calcul très complexes. La régression logistique est aussi efficace pour prétraiter les données (ex : en les triant) afin de faciliter les calculs par la suite.

Par ailleurs, la régression logistique est un modèle robuste qui permet de faire face à la non-normalité des prédicteurs, et ne nécessite pas que les erreurs soient distribuées normalement. Les coefficients de la régression logistique sont facilement interprétables les uns par rapport aux autres à partir d'une valeur de référence (*intercept*). Enfin, ce modèle permet l'évaluation de l'ajustement et la performance du modèle à travers différents outils (ex : R^2 de McFadden, test de Hosmer-Lemeshow, etc.). La régression logistique utilise la fonction *glm* dans R permettant de prédire le résultat d'une observation (y) à partir d'une variable prédictive (x) telle que :

$$p = \frac{e^y}{1+e^y} = \frac{1}{1+e^{-y}}$$

Où : $y = \beta_0 + \beta_1 x$ et β_0 et β_1 les coefficients de régression, p est la probabilité que l'événement se produise étant donné x . β_1 positif indique que l'augmentation de x sera associée à l'augmentation de p . Inversement, un β_1 négatif indique que l'augmentation de x sera associée à une diminution de p ([Multinomial Logistic Regression | R Data Analysis Examples, 2024](#)). En généralisant, on peut démontrer que :

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n$$

On notera que le logarithme de la quantité $\log[p/(1-p)]$ est appelé logarithme de l'impair, également connu sous le nom *logit*, ce qui sera utile par après. La cote ($exp(y)$) désigne la probabilité que l'événement se produise et on cherche alors le rapport entre les succès et les non-succès. La régression logistique permet d'estimer la probabilité (p) d'appartenance à une catégorie sans indiquer à quelle catégorie précise la variable donnée appartient. Il faut alors définir un seuil de probabilité (généralement $p=0,5$ par défaut) à partir duquel variable passe d'une catégorie à l'autre.

Lors de la régression, la valeur de référence (*intercept*) est calculée lorsque toutes les variables du modèle sont évaluées à zéro. Cette valeur se calcule en prenant le logarithme des probabilités de

l'ensemble de la population d'intérêt d'appartenir à la variable d'intérêt principal (la multimodalité) sans variables prédictives dans le modèle. Par défaut, le logiciel R prend comme référence la première valeur implémentée dans le modèle de régression. Les autres variables ou catégories sont comparées ensuite à cette valeur de référence via des valeurs d'estimation (*estimates*). Si la valeur d'estimation est positive, alors la variable considérée a une probabilité plus élevée d'obtenir un succès vis-à-vis de la variable d'intérêt principal. Une valeur négative indique que la probabilité d'obtenir un succès est plus faible, soit dans notre cas d'être multimodal.

Chaque variable prédictive est associée à une *p-value* qui indique l'importance de la variable dans le modèle (si elle est significative ou non). On notera par la suite une légende associée au niveau de pertinence de la variable par rapport à la variable d'intérêt principal : * quand *p-value* < 0,01, ** quand *p-value* < 0,001, et *** quand *p-value* < 0,0001.

Tableau 14 : Exemple de régression logistique "simple" sur la variable "WHYTO" (ou "PURPOSE")

Variable	Estimate	Std. Error	z value	p-value	Pertinence
WHYTO1 (Référence)	-4,58	0,50	-9,12	<0,01	***
WHYTO2	0,76	0,52	1,47	0,14	
WHYTO3	2,77	0,51	5,42	<0,01	***
WHYTO4	0,14	0,52	0,27	0,79	
WHYTO5	3,87	0,56	6,93	<0,01	***

De informations complémentaires peuvent être obtenues sur le modèle de régression logistique utilisé en fonction des données d'entrée :

- La déviance nulle ou déviance du même modèle sans variables prédictives calculé en comparant le modèle ajusté à un modèle avec seulement un terme d'origine (intercept seul),
- La déviance résiduelle ou déviance du modèle ajusté, qui mesure la façon dont le modèle s'ajuste aux données après avoir pris en compte les variables prédictives.
- L'AIC (*Akaike Information Criterion*) qui mesure de la qualité de l'ajustement du modèle, ajustée en fonction du nombre de paramètres du modèle (les modèles les mieux adaptés ont un AIC plus faibles).
- Le nombre d'itérations de Fisher Scoring qui indique le nombre d'itérations effectuées par l'algorithme de Fisher Scoring pour estimer les paramètres du modèle.

Tableau 15 : Exemple de régression logistique "simple" sur la variable "WHYTO" (ou "PURPOSE")

VARIABLES	Valeurs
Intercept	-3,55
Degrés de liberté (Df)	10590
Null Deviance	2701,52
Residual Deviance	2322,12
AIC	2703

Le modèle de régression logistique peut être appliqué à plusieurs variables en même temps mais également à une seule variable à la fois. Dans notre cas, il s'agit de comparer l'impact de chaque variable sur la multimodalité par rapport aux autres variables (logistique générale ou *Logistic_Tot* dans le modèle sur R). Ensuite, une régression peut être envisagée par variable vis-à-vis de la multimodalité afin de montrer l'impact de chaque catégorie associée à chacune des variables (ex : régression simple entre l'âge *R_AGE* et la multimodalité *Multi*).

3.4.2 Multicolinéarité et lien entre variables

La création d'intervalles (distance, temps de trajet et âge) visait à simplifier le travail d'analyse en regroupant les participants dans de grandes catégories plutôt que de conserver des variables avec beaucoup de catégories différentes (variables considérées comme quasi continue comme l'âge par exemple). Ce travail induit de fait une forte colinéarité entre les variables, alors liées deux à deux (ex : *Age_Intervals* liée à *R_AGE*) ce qui peut altérer la qualité du modèle.

La multicolinéarité des variables d'un modèle donné peut être évaluée grâce au facteur d'inflation de la variance généralisée (ou *GVIF* par après). Ce facteur quantifie à quel point la variance du coefficient estimé est plus grande par rapport à ce qu'elle serait si ce prédicteur n'était pas corrélé avec les autres prédicteurs du modèle. Le diagnostic VIF est utile pour détecter la multicolinéarité entre plusieurs variables mais il ne fournit pas d'informations sur la direction ou la nature des relations entre les variables prédictives.

Dans la pratique, on cherchera à calculer la racine carrée de la valeur *GVIF* élevée à la puissance 1 divisée par deux fois les degrés de liberté associés à chaque variable prédictive (troisième colonne du tableau ci-dessous). Cette valeur sera notée valeur *VIF* par la suite. En général, une valeur *VIF* supérieure à 10 indique une multicolinéarité élevée et des valeurs supérieures à 5 sont souvent considérées comme préoccupantes. Au contraire, les valeurs proches de 1 suggèrent peu ou pas de problèmes de multicolinéarité. On obtient pour les variables retenues les valeurs suivantes :

Tableau 16 : Colinéarité entre les variables

Variables	GVIF	Df	$GVIF^{1/(2*Df)}$ ou VIF
ADULTS	130,67	3	2,25
Age_Intervals	102,69	3	2,16
Distance_Intervals	20,11	3	1,65
DRIVER	6,06	1	2,46
DRIVERS	48,32	4	1,62
EDUC	8,99	6	1,20
HHFAMINC	1,58	3	1,08
HOUSEVEH	5,13	4	1,23
LIF_CYC	83,05	5	1,56
PRMACT	12,36	5	1,29
R_AGE	14,60	1	3,82
R_RACE	1,24	5	1,02
R_SEX	1,08	2	1,02
TRAVDAY	1,31	6	1,02
Travel_time_Intervals	11,25	3	1,50
TRPMILES	14,46	1	3,80
TRVLCMIN	7,55	1	2,75
URBANSIZE	2,96	3	1,20
URBRUR	2,78	1	1,67
WHYTO	3,67	9	1,07

Certaines variables sont naturellement corrélées deux à deux car elles sont établies les unes par rapport aux autres et cela explique une colinéarité plus grande (*GVIF* plus élevé) :

- *Age_Intervals* basé sur *R_AGE*
- *Distance_Intervals* basé sur *TRVLCMIN*

- *Travel_time_Intervals* basé sur *TRVLCMIN*

La colinéarité des variables par rapport aux autres ne dépasse pour le moment jamais la valeur limite de 4 donc les variables mise en valeur dans le précédent tableau ne sont pas trop liées entre elles. En revanche certaines valeurs s'en rapprochent dangereusement (ex : *R_AGE*, *TRPMILES* et *TRVLCMIN*), ce qui démontre un certain lien quand même entre les variables. Cela est lié au travail de création des intervalles qui viennent en quelque sorte dupliquer les variables existantes.

Sachant cela, l'ensemble des variables sera conservé pour le moment mais le *VIF* sera recalculé à la fin du processus de tri des variables (voir *Régression de type III* par après) afin de ne conserver que les variables non corrélées.

3.4.3 Sélection des variables d'intérêt

Le but maintenant est de ne conserver que les variables les plus significatives de la régression logistique générale afin de renforcer la pertinence du modèle. On effectue pour cela une analyse de la variance (ou régression de type III) qui permet de tester la différence entre les moyennes de plusieurs groupes, en supposant que les données suivent une distribution normale et que les variances sont égales. En d'autres termes, cette fonction permet de déterminer si les modifications apportées aux variables indépendantes ont un effet significatif sur la variable dépendante (dans notre cas la multimodalité). Cela passe une comparaison de la variance du modèle avec variable considérée à celle du modèle sans cette variable en ajoutant les autres variables du modèle une par une selon l'ordre dans lequel elles apparaissent dans le modèle.

La régression de type III effectue différents tests à la suite pour chaque variable d'intérêt : un modèle avec seulement l'intercepte, puis un modèle avec l'intercepte et le premier prédicteur et enfin un modèle avec l'intercepte, le premier prédicteur, et le deuxième prédicteur. Cette méthode est particulièrement utile lorsque les prédicteurs sont corrélés ou dans les cas de facteurs factoriels et d'interaction, où il est important de comprendre l'impact de chaque prédicteur indépendamment des autres.

On précise que la variante utilisée de cette fonction est le type III car les termes du modèle ne sont pas hiérarchiquement ordonnés et également pour avoir plus de contrôle sur les effets des autres variables sur la variable considérée. Cela signifie que chaque terme est testé pour son effet unique de façon indépendante, en supposant que tous les autres termes sont également inclus dans le modèle. Concernant données et valeurs retenues dans l'étude, les variables sexe et âge seront considérés comme des variables d'intérêt qui seront conservées quoi qu'il arrive par la suite. Voici les résultats de la régression de type III effectuée sur l'ensemble du modèle (première itération) :

Tableau 17 : Première itération de la régression de type III sur l'ensemble des variables d'intérêt (exemple)

VARIABLES	LR Chisq	Df	p-value	Pertinence
ADULTS	5,54	3	0,14	
Age_Intervals	0,99	3	0,80	
Distance_Intervals	31,30	3	<0,01	***
DRIVER	9,15	1	<0,01	**
DRIVERS	11,80	4	0,02	*
EDUC	9,20	6	0,16	
HHFAMINC	5,42	3	0,14	
HOUSEVEH	48,19	4	<0,01	***
LIF_CYC	8,55	5	0,13	
PRMACT	6,06	5	0,30	
R_AGE	0,89	1	0,35	
R_RACE	11,17	5	0,05	*
R_SEX	7,33	2	0,03	*
TRAVDAY	23,29	6	<0,01	***
Travel_time_Intervals	36,98	3	<0,01	***
TRPMILES	3,54	1	0,06	.
TRVLCMIN	3,56	1	0,06	.
URBANSIZE	14,82	3	<0,01	**
URBRUR	0,51	1	0,48	
WHYTO	125,05	4	<0,01	***

Où *Df* correspond aux degrés de liberté pour chaque terme du modèle (*degrees of freedom*). *LR Chisq* (ou *la déviance*) mesure de l'écart entre le modèle complet et le modèle sans le terme spécifié et plus la déviance est élevée, plus le terme est important. La *p-value* associée au test de rapport de vraisemblance (ou Chi-carré), indique elle la probabilité d'observer des résultats aussi extrêmes que ceux observés, sous l'hypothèse nulle que le terme ajouté n'a pas d'effet sur la variable de réponse.

Les étoiles à côté des p-valeurs indiquent le niveau de significativité statistique et plus il y a d'étoiles, plus la p-value est faible et plus le terme est significatif. On notera la présence de valeurs limites de *p-value* pertinence (marquées par un point et au niveau de la pertinence) qui indiquent que la variable donnée n'est pas suffisamment significative et qu'elle doit être exclue du modèle.

A chaque itération de la fonction de type III, la variable la plus significative (*p-value* la plus élevée) n'est pas conservée pour les prochaines itérations (ici *Age_Intervals*), car elle constitue la variable la moins significative vis-à-vis de la multimodalité (variable d'intérêt principal). L'opération est alors répétée jusqu'à ce qu'il n'y ait plus des variables significatives à la fin (ou une pertinence suffisamment grande):

Tableau 18 : Dernière itération de la régression de type III sur l'ensemble des variables significatives conservées

Variables	LR Chisq	Df	p-value	Pertinence
Distance_Intervals	31,82	3	<0,01	***
DRIVER	7,58	1	<0,01	**
HOUSEVEH	70,81	4	<0,01	***
R_AGE	17,67	1	<0,01	***
R_SEX	10,55	2	<0,01	**
TRAVDAY	22,33	6	<0,01	<0,01
Travel_time_Intervals	99,01	3	<0,01	***
TRPMILES	5,07	1	0,02	*
URBANSIZE	18,33	3	<0,01	***
WHYTO	143,52	4	<0,01	***

Au terme du processus de sélection des variables via la fonction de type III, seules 10 variables sont considérées comme pertinentes au terme de 11 itérations. Le *VIF* est alors recalculé pour revérifier la multi colinéarité éventuelle entre les variables restantes :

Tableau 19 : Calcul du facteur *VIF* pour les variables significatives

Variables	GVI	Df	$GVI^{1/(2*Df)}$ ou <i>VIF</i>
DRIVER	1,89	1	1,38
Distance_Intervals	19,06	3	1,63
HOUSEVEH	1,21	4	1,02
R_AGE	1,68	1	1,29
R_SEX	1,02	2	1,01
TRAVDAY	1,21	6	1,02
Travel_time_Intervals	3,80	3	1,25
TRPMILES	13,89	1	3,73
URBANSIZE	1,09	3	1,02
WHYTO	1,98	4	1,09

Comme on pouvait s'y attendre, la distance (*TRPMILES*) et les intervalles de distance (*Distance_Intervals*) sont assez corrélées. En revanche, le seuil de 4 pour la valeur de *VIF* n'a pas été dépassé. Pour la suite de l'analyse, on privilégiera les intervalles de distance pour faciliter l'interprétation et le lien avec la littérature.

On se retrouve ainsi avec 9 variables d'intérêt qui seront confrontées par la suite dans la partie résultats et analyse, en lien avec la littérature. Pour chaque variable conservée, une régression logistique simple peut être effectuée sur la variable considérée pour classer par ordre d'importance les sous-catégories associées. Sur le modèle régression effectuée sur l'ensemble des variables, on cherchera donc à lier chaque variable à la variable de référence *Multimodalité* (ou *Multi*) selon un modèle *glm* (voir formule de la régression logistique plus haut).

3.5 Pertinence et limites du modèle

Les précédents résultats n'ont de pertinence que si le modèle en lui-même est pertinent. Un certain nombre de tests peuvent être conduits pour cela, c'est ce que nous allons voir maintenant.

3.5.1 ROC

L'analyse ROC (pour *Receiver Operating Characteristic*) permet d'évaluer l'exactitude des prévisions d'un modèle en traçant la sensibilité par rapport au taux de faux positifs d'un test de classification. Elle prend des valeurs entre 0 et 1, où une valeur de 1 indique une classification parfaite (c'est-à-dire que le modèle distingue parfaitement les classes), et une valeur de 0,5 indique un modèle sans pouvoir discriminant, équivalent à une supposition aléatoire (Bartlett, 2014b). Le ROC a été calculé au début et à la fin du processus de sélections des variables et on obtient alors les résultats suivants :

Tableau 20 - Résultats du test ROC sur le modèle de régression logistique

Test ROC	Aire sous la courbe
ROC pré type III	0,90
ROC post type III	0,89

Ainsi, près de 90% de cellules ont bien été identifiées par rapport à la variable d'intérêt principale *Multimodalité* pour ce modèle. Cela signifie que le modèle de régression logistique choisi (*Logistic_reg_Tot* dans le code R en annexe) a un fort pouvoir discriminatif pour distinguer les classes de votre variable de réponse en ce qui concerne la multimodalité car l'air sous la courbe est supérieure à 0,8, ou 80% (Çorbacioğlu & Aksel, 2023). En d'autres termes, il est capable de séparer efficacement les deux sous-catégories pour chaque variables considérées vis-à-vis de la multimodalité des personnes interrogées. Il faut cependant garder à l'esprit que ce résultat ne fonctionne que dans le contexte de cette analyse et n'est donc pas directement comparable à un modèle de référence.

3.5.2 Hosmer-Lemeshow

La pertinence du modèle peut également être vérifiée via le test de Hosmer-Lemeshow qui contrôle la qualité de l'ajustement et de l'étalement de la régression logistique. Dit autrement, ce test regarde si les taux observés correspondent aux événements attendus et permet de déterminer dans quelle mesure le modèle s'ajuste aux données (Nattino et al., 2020). Le test calcul notamment les probabilités prédites à partir du modèle de régression logistique (*fitted(logistic_model)* dans le code) pour les 10 variables finales considérées. Une *p-value* significative inférieure à 0,5 suggère que le modèle n'est pas bien ajusté aux données. Au contraire, une *p-value* qui tend vers 1 indique un bon ajustement des données, ou qu'il n'y a pas de différence significative entre les valeurs observées et prédites. Voici les résultats obtenus pour le modèle utilisé :

Tableau 21 : Résultats du test Hosmer-Lemeshow sur le modèle de régression logistique

X-squared	Degrés de liberté (Df)	p-value
6,27	8	0,62

Une *p-value* de 0,62 indique un bon ajustement des données car est bien au-dessus de 0,05 (Zimmerman & Kramer, 2008). En revanche, ce résultat indique que le modèle pourrait être davantage amélioré notamment afin de tendre vers 1. Le modèle fonctionne davantage lorsqu'il n'y a que des variables significatives mais comme on l'a vu, c'est le cas ici. La variable liée à la distance (*TRPMILES*) pourrait être supprimée car on a déjà les intervalles de distance (*Distance_Intervals*), ce qui affinerait légèrement le modèle.

Une autre méthode serait d'améliorer le nettoyage et la restructuration des données en examinant les valeurs trop extrêmes ou encore en limitant le nombre de degrés de liberté pour chaque variable observée. Le problème dans ce dernier est qu'il faut également conserver une certaine cohérence pour pouvoir ensuite interpréter correctement ces données par la suite. On évitera donc de trop concaténer les différentes sous-catégories de chaque variable étudiée pour ces raisons. Enfin, ce résultat peut être imputé au nombre assez faible de personnes multimodales par rapport à la taille de l'échantillon (296 sur 10591 soit environ 2,79%).

3.5.3 MacFadden

Enfin, le test de MacFadden (ou *MacFadden Rho Squared*) permet de tester la qualité du modèle, c'est-à-dire que les données sont bien pertinentes dans le modèle (Bartlett, 2014). On précise que les résultats de ce test ne sont pas comparables à ceux d'un autre modèle car ils sont fortement liés au jeu de données utilisé. On précise également que l'on considère ici la version *Market Share* et non *Equal share*. La valeur d'intercept utilisée pour le test est la même que celle du modèle de régression logistique utilisé jusque-là, soit la première variable implémentée dans le modèle (*R_SEX* et plus spécifiquement la catégorie *Female* dans notre cas).

Tableau 22 : Données d'entrées du test de MacFadden Rho Squared

Variable	Value
Intercept	-3,55
Degrés de liberté (Df)	10591
Null Deviance	2701
Residual Deviance	2701
AIC	2703

On calcule ensuite le logarithme de la vraisemblance (ou *likelihood*) de l'intercept seul que l'on compare à la vraisemblance de l'ensemble du modèle. En d'autres termes, il faut calculer la vraisemblance du modèle en entier avec toutes les variables d'intérêt (*log Lik.' (Full)*) ajustée sur la vraisemblance du modèle indépendant ou avec l'intercept seul (*log Lik.' (null)*). La valeur finale du test s'obtient en soustrayant à 1 la valeur du rapport entre les deux vraisemblances (formule : $1 - \log \text{Lik.}(\text{Full}) / \log \text{Lik.}(\text{null})$). On obtient alors les résultats suivants :

Tableau 23 : Résultats du test de MacFadden Rho Squared

Vraisemblance (log Likelihood)	Valeur	Degrés de liberté (Df)
log Lik. (Full)	-968,59	29
log Lik. (null)	-1350,74	1
log Lik. final	0,28	29

La valeur finale doit être comprise entre 0 et 1. Une valeur au-dessus de 0,4 et tendant vers 1 indique que le modèle a une forte valeur explicative, ce qui témoigne de la solidité du modèle. Au contraire, une valeur en dessous de 0,2 et tendant vers 0 indique que le modèle n'a pas de valeur explicative au-delà de l'intercept seul (ou la valeur de référence) et qu'il n'est pas solide (What's the Best R-Squared for Logistic Regression?, 2024, Hensher & Stopher, 2021).

Ici, la valeur finale est intermédiaire (entre 0,2 et 0,4) donc cela signifie que le modèle a une certaine valeur explicative au-delà de l'intercept, dans le sens qu'il y contribue mais qu'il doit être amélioré. Les pistes d'amélioration pour le test de Hosmer Lemeshow pourraient également augmenter les résultats pour le test de MacFadden. Des études plus poussées sur l'impact du contexte encadrant les données et de la complexité du modèle doivent être menées pour expliquer plus en détail les précédents résultats.

3.5.4 Table de classification

La table de classification donne des indications sur les performances du modèle de régression logistique sélectionné ([datasciencebeginners, 2020](#)). Elle est construite à partir de valeurs réelles et les valeurs prédites et permet de donner différentes informations clés sur le modèle : l'exactitude, le taux d'erreur de classification, le taux de vrais positifs, également connu sous le nom de rappel, le taux de vrais négatifs et la précision.

Tableau 24 : Table de classification

1 : multimodal, 0 : non multimodal		Valeurs réelles	
		1	0
Valeurs prédites	1	15 (FP)	21 (TP)
	0	10280 (TN)	275 (FN)

Avec :

- TP (*True positive*) : correspond au nombre de vrais positifs, où le modèle a correctement prédit la classe positive.
- TN (*True negative*) : correspond au nombre de vrais négatifs, où le modèle a correctement prédit la classe négative.
- FP (*False positive*) : correspond au nombre de faux positifs, où le modèle a mal prédit la classe positive.
- FN (*False negative*) : correspond au nombre de faux négatifs, où le modèle a mal prédit la classe négative.

Plusieurs indicateurs peuvent être alors calculés à partir des valeurs de la table de classification :

Tableau 25 : Résultats associés à la table de classification

Variables	Formule	Valeur (%)
Précision (<i>Accuracy</i>)	$(TP+TN)/(TN + FP + FN + TP)$	97,26
Taux de misclassification	$(FP+FN)/(TN + FP + FN + TP)$	2,74
Taux de réels positifs ou <i>Recall</i> (TPR)	$TP/(FN + TP)$	7,09
Taux de réels négatifs (TNR)	$TN/(TN + FP)$	99,85
Valeur prédictive négative (<i>Precision</i>)	$TP/(FP + TP)$	58,33
F-Score	$(2 * Précision * Recall / (Précision + Recall))/100$	12,65

La précision (*Accuracy*) correspond au rapport du nombre total de prédictions correctes (à la fois vraies positives et vraies négatives) sur toutes les prédictions effectuées. Une précision de 97,26 % est relativement élevée, indiquant que le modèle fonctionne bien dans l'ensemble mais il est à coupler

avec les autres indicateurs pour plus de détail. Cela est confirmé par un taux de misclassification très faible (inférieur à 3%) qui lui indique la proportion de valeurs prédites qui sont fausses ([datasciencebeginners, 2020, Saito & Rehmsmeier, 2015](#)).

Mais d'un autre côté, le TPR ou taux de rappel (*Recall*), qui mesure la proportion de vrais positifs correctement identifiés, a une valeur faible (7,09%). Cela indique que le modèle fonctionne mal pour identifier la classe positive lorsqu'elle est vraiment positive. Autrement dit, le modèle a du mal à identifier les personnes multimodales quand elles le sont vraiment. Au contraire, le TNR, qui mesure la proportion de vrais négatifs correctement identifiés, a une valeur élevée (99,85 %). Cela indique que les non multimodaux, très nombreux, sont correctement identifiés comme non multimodaux.

Le *F-Score* est un moyen harmonique de rappel et de précision dont la valeur se situe entre 0 et 1. La valeur de 1 représente une grande précision et un rappel parfait. La valeur 0 représente au contraire le pire des cas où la précision est très faible. Ici, un faible *F-Score* (12,65%) indique une mauvaise performance d'un modèle de classification, en particulier en termes d'équilibre entre la précision et le TPR (*Recall*) car le nombre élevé de faux positifs par rapport aux vrais positifs impacte directement le *F-Score*.

Ces résultats peuvent s'expliquer par le très faible nombre de personnes multimodales (296) par rapport aux personnes non multimodales (10295), ce qui crée un déséquilibre qui aurait tendance à privilégier la prédiction de la classe la plus répandue (ici les non multimodaux). Des stratégies de rééquilibrage basées sur la notion de coût permettraient de mieux équilibrer les classes. Le seuil de classification pour la multimodalité peut aussi être revu afin d'avoir un plus grand nombre de personnes multimodales (cf. parties 3.2.2).

3.5.5 Limites de la régression logistique

Le modèle de régression logistique est relativement facile à mettre en place et plutôt rapide par rapport à d'autres modèles. En revanche, il y a un certain nombre de limitations pour les résultats binaires. La relation entre les prédicteurs et la probabilité du résultat est généralement non linéaire, ce qui implique des relations complexes entre les prédicteurs et le résultat. Des modèles plus complexes tels que les arbres de décision, les forêts aléatoires ou les réseaux neuronaux peuvent être plus performants et prennent mieux en compte cette non-linéarité ([Couronné et al., 2018](#)).

De plus, comme on l'a vu, la gestion des déséquilibres des données (cf. partie précédente) est mal prise en compte dans le modèle actuel, car il y a beaucoup plus de personnes non multimodales que multimodales. Les hypothèses définissant la multimodalité peuvent être revues pour espérer augmenter la part de personnes multimodales (ex : considération de la marche comme mode propre), mais dans les faits, cette proportion restera toujours très faible devant celle des non multimodaux. Des tests d'erreurs distribués pourraient être effectués en complément de cette analyse afin de montrer ces déséquilibres.

Face à ce problème, on aurait pu classer les données selon un coût (ou un poids) plus élevé pour les multimodaux pour qu'ils soient mieux représentés au sein de la population globale. De même, la technique du *Resampling* aurait pu être appliquée aux personnes non multimodales pour donner plus de poids aux multimodaux ([Resampling to Properly Handle Imbalanced Datasets in Machine Learning - Comet, 2024](#)).

Enfin, le logiciel ne permet pas de comparer directement les différentes catégories de variables les unes par rapport aux autres. Le résultat de la régression donne une valeur correspondant au logarithme des cotes (ou *Odds*) associées à chaque catégorie. Il faut ensuite transformer manuellement ces valeurs en cotes pour pouvoir les comparer (en calculant l'exponentiel de cette du log notamment) et cela complexifie davantage l'analyse par la suite.

3.5.6 Limites du logiciel R

Le logiciel R a été conçu à la base par des statisticiens afin de conduire des études statistiques avancées (possibilité de télécharger de nombreux packages dédiés). Ce logiciel, gratuit et open source, a donc naturellement été choisi pour analyser les données du NHTS, mais d'autres logiciels auraient tout aussi bien pu convenir (ex : programme Python).

La principale limitation de R est sa puissance de calcul relativement faible comparée à d'autres logiciels, ce qui est notamment dû au fait que le logiciel n'est pas compilable ([R++, the Next Step_C. Genolinia, J. Falcoub, et R. Tournierc \(2010\).pdf, s. d.](#)). La mémoire est rapidement utilisée au maximum et ce même sur un échantillon de valeurs assez faibles, comme dans notre cas (au maximum 30 074 déplacements à traiter) alors que le logiciel est capable en théorie de traiter des millions d'opérations en même temps. A titre d'exemple, la première grosse opération du programme consistait à détecter une séquence multimodale pour chaque personne interrogée en se basant sur le tableaux des déplacements (*data_TRIP*).

Cette opération a pris plusieurs heures à s'effectuer avec un ordinateur assez puissant, même s'il faut reconnaître qu'une connaissance plus approfondie du logiciel aurait permis d'optimiser davantage le code (et donc de diminuer le temps de réponse). Les objets sont directement stockés dans la mémoire physique du logiciel à un seul endroit alors que d'autres logiciels (ex : Python) stockent l'information à différents endroits de la mémoire, ce qui limite la charge lors des calculs ([R Advantages and Disadvantages - Javatpoint, 2024](#)).

Une autre limitation du logiciel, est le fait que l'on ne doit pas laisser de cellules vides dans nos jeux de données lorsque l'on veut en faire l'analyse, en particulier lors d'une régression logistique ([Logit Regression | R Data Analysis Examples, 2024](#)). Les cellules vides (*NA* dans les tableaux) peuvent augmenter l'instabilité du modèle global et ainsi l'empêcher de fonctionner. De plus, la présence de cellules vides nuit à l'analyse par après car une partie des usagers ou des déplacements ne peuvent pas être décrits.

Enfin, une autre limitation du logiciel est la nécessité de factoriser chaque variable avant d'effectuer la régression logistique globale. Cela implique d'avoir deux fois la même variable dans notre jeu de données (ex : *R_RACE* et *RACE*) : une version sous forme de facteur ou niveaux (nombre entier utilisable par le modèle *glm*) et une version avec la vraie valeur sous format texte correspondante (ex : chiffre 3 dans *R_RACE* associé à *Hispanic* dans *RACE*). Le logiciel a en effet du mal à factoriser directement des chaînes de caractères comme du texte et au contraire préfère traiter des valeurs numériques directement.

3.6 Résultats et discussion

3.6.1 Proportion de multimodaux

Les règles définissant le caractère multimodal d'un trajet ou d'un déplacement ont été précédemment établies de façon arbitraire mais en cherchant à se rapprocher au mieux de la réalité (voir partie 3.2.2). Dans le premier cas (tableaux des trajets), il s'agissait de regarder la séquence des déplacements d'un usager sur la journée afin de mettre en lumière une séquence multimodale dans son parcours. Ceci permettant alors de qualifier la personne de multimodale ou non. Dans le second cas (tableaux des déplacements), il s'agissait de regarder une partie plus fine de la séquence pour déceler la multimodalité d'une personne mais sur deux déplacements consécutifs. La personne interrogée est considérée comme multimodale si son trajet ou son déplacement constitue une séquence multimodale. Voici les résultats obtenus rapportés à la population totale de l'étude :

Tableau 26 : Proportion de trajets et de déplacements multimodaux

Variables	Cas du trajet	Cas du déplacement
Nombre d'occurrences multimodales	296 (2,79%)	952 (3,06%)
Nombre d'occurrences non multimodales	10295 (97.21%)	30122 (96,94%)
Total	10591 (100%)	31074 (100%)

Comme on pouvait s'y attendre, la proportion de personnes multimodales est très faible, que ce soit sur le trajet total ou sur le déplacement. En effet, la place de la voiture individuelle reste prépondérante aux Etats-Unis et ce même au cœur des villes, notamment par rapport à d'autres régions du monde (ex : en Europe). Ces scores faibles peuvent cependant être améliorés en redéfinissant la multimodalité dans chaque cas.

Pour le trajet, une des principales limitations a été de ne pas considérer la marche car elle était mal définie dans les données de base et surtout non comptabilisée comme un déplacement en soi. Les données fournies par le NHTS ne permettaient pas d'avoir beaucoup d'informations sur la marche par rapport aux autres modes de transport (temps de trajet, distance parcourue, but du déplacement, etc.), ce qui ne permet pas d'inclure ce mode dans la séquence des déplacements. Le caractère multimodal de la marche dépend de chaque situation car une personne qui marche du parking au quai du train peut être considérée comme tout autant multimodale qu'une autre personne qui marcherait pendant 30 minutes pour se rendre à la gare depuis chez elle.

On remarque également que la proportion de multimodaux est légèrement plus importante dans le cas du déplacement que dans le cas du trajet. Cela est dû, d'une part, à la définition plus contraignante dans le cas du trajet pour définir la multimodalité (plus de critères). D'autre part, car les personnes multimodales comptabilisent en moyenne 3,86 déplacements au total là où les non multimodaux comptabilisent 2,91 déplacements seulement en moyenne (soit une augmentation de 32% en moyenne). Dans le cas des déplacements, la multimodalité est évaluée sur deux déplacements à la suite et un nombre moyen de déplacements plus important chez les multimodaux permet d'atteindre une proportion finale de personnes multimodales plus élevée que dans le cas des trajets.

3.6.2 Facteurs influençant la multimodalité

Le travail précédent a conduit à la sélection de 9 des 20 variables sélectionnées au départ, avec des variables quasi continue (ex : l'âge) ou d'autres limitées à quelques catégories (ex : 3 pour le sexe).

Parmi les variables restantes, ont été conservées des variables d'intérêt majeur (l'âge et le sexe) car elles sont considérées comme cruciales pour différencier les personnes interrogées dans l'étude. Deux variables créées de toute pièce ont été privilégiées aux variables existantes liées afin de faciliter l'analyse par la suite (intervalles de distance et de temps de trajet à la place de la distance et du temps absolu pour chaque trajet). Voici finalement les variables retenues :

Tableau 27 : Variables finales et pertinence

Variables	LR Chisq	Df	p-value	Pertinence
Distance_Intervals	31,82	3	<0,01	***
DRIVER	7,58	1	<0,01	**
HOUSEVEH	70,81	4	<0,01	***
R_AGE	17,67	1	<0,01	***
R_SEX	10,55	2	0,01	**
TRAVDAY	22,33	6	<0,01	**
Travel_time_Intervals	99,01	3	<0,01	***
URBANSIZE	18,33	3	<0,01	***
WHYTO	143,52	4	<0,01	***

La pertinence de chaque variable a été évaluée sur l'ensemble des variables de l'étude. Il reste maintenant à évaluer la pertinence de chaque catégorie associée aux différentes variables afin de déterminer les facteurs qui impactent le plus la multimodalité des personnes. Pour cela, une régression logistique dite *simple* est conduite entre chaque variable retenue et la variable d'intérêt principale (la multimodalité). Avec la même logique que pour la régression générale, les catégories seront classifiées par ordre d'importance au sein de chaque variable et les résultats seront confrontés à la littérature.

On précise que la valeur de référence (*Intercept*) sélectionnée par défaut par le logiciel R pour chaque variable sera la première catégorie par ordre alphabétique. Cette valeur peut être changée à tout moment grâce à la fonction *Relevel* dans le programme R afin de privilégier les valeurs limites comme valeur de référence. Les autres catégories seront donc comparées à cette catégorie de référence.

Tableau 28 : Résultats de la régression logistique générale et cotes relatives entre les catégories pour chaque variable

Variable	Catégorie	Estimate	Cote	Effet sur les cotes (%)	p-value	Pertinence
Variables socio-économiques						
DRIVER	Driver	0,00	-	-	-	-
	Not a driver	0,52	1,69	68,58	0,01	**
HOUSEVEH	0 véhicule	0,00	-	-	-	-
	1 véhicule	-1,32	0,27	-73,40	<0,01	***
	2 véhicules	-1,94	0,14	-85,61	<0,01	***
	3 véhicules	-1,92	0,15	-85,30	<0,01	***
	4+ véhicules	-2,04	0,13	-86,93	<0,01	***
R_AGE	R_AGE	-0,02	0,98	-1,95	<0,01	***
R_SEX	Female (Référence générale)	-5,30	0,01		<0,01	***
	Male	0,39	1,47	47,02	<0,01	**
	Other (sex)	0,80	2,21	121,48	0,08	.
Variables liés au contexte						
Distance Intervals	(0,10[0,00	-	-	-	-
	(10,25[3,00	20,17	1917,47	<0,01	***

	(25,50[1,28	3,60	260,36	<0,01	***
	(50,5000[2,08	8,03	702,59	<0,01	***
TRAVDAY	Sunday	0,00	-	-	-	-
	Monday	1,48	4,37	337,13	<0,01	***
	Tuesday	1,19	3,27	227,48	0,01	**
	Wednesday	1,56	4,78	377,56	<0,01	***
	Thursday	1,24	3,45	245,47	<0,01	**
	Friday	1,20	3,32	231,58	0,01	**
	Saturday	1,44	4,22	321,75	<0,01	**
Travel Time Intervals	(0,30[minutes (Référence locale)	0,00	-	-	-	-
	(30,60[minutes	0,09	1,09	9,28	0,60	
	(60,120[minutes	-0,83	0,44	-56,22	<0,01	***
	(120,1140[minutes [-1,28	0,28	-72,19	<0,01	***
URBANSIZE	50,000-499,000	0,00	-	-	-	-
	500,000-999,999	-0,47	0,63	-37,26	0,16	
	1,000,000 or more	0,46	1,59	58,75	0,01	*
	Other (Density)	0,54	1,71	70,82	0,01	**
WHYTO	Regular activities at home	0,00	-	-	-	-
	Attend school as a student	0,62	1,85	85,07	0,26	
	Change type of transportation	1,95	7,05	605,24	<0,01	***
	Other Mandatory	0,11	1,12	11,69	0,84	

Possession du permis de conduire (DRIVER) :

Le permis de conduire est accessible dès 16 ans aux Etats-Unis et ce avant même d’atteindre la majorité (à 21 ans) ni même d’être rentré dans la vie active (à cet âge les américains sont toujours au lycée ou *High Schools*). Cette accessibilité assez précoce à un véhicule (par rapport au permis à 18 ans en Europe par exemple) favorise la démocratisation de la voiture individuelle au détriment des autres modes. A l’inverse, il est attendu que l’absence d’un permis de conduire favorise grandement l’utilisation de modes alternatifs et donc augmente la multimodalité des usagers.

Tableau 29 : Résultats de la régression logistique générale et cotes relatives (DRIVER), extrait du tableau 28

Catégorie	Estimate	Exp(estimate)	Effet sur les cotes (%)	p-value	Pertinence
Driver (Référence locale)	0,00	-	-	-	-
Not a driver	0,52	1,69	68,58	0,01	**

Les résultats sont catégoriques : le fait de ne pas posséder un permis de conduire augmente de plus de 68% les chances d’être multimodal, parmi les personnes considérées comme multimodales. Cela fait directement écho aux résultats d’[Heinen & Kiron Chatterjee \(2015\)](#) ou encore de [Sungyop Kim & al. \(2007\)](#), même si on l’a vu, cette tendance tend à diminuer avec le temps ([Heinen & Mattioli, 2019](#)).

A l’inverse, comme l’avaient prévus [Blumenberg & Pierce \(2013\)](#) et [Eva Heinen & Kiron Chatterjee \(2015\)](#), le fait d’avoir le permis de conduire diminue, et par extension une voiture, réduit drastiquement

les chances d'utiliser un autre modes de transport alternatif, même si l'on peut toujours être multimodal (ex : aller à la gare en voiture pour aller travailler en centre-ville).

D'un autre côté, [Bai & Kattan \(2014\)](#) ou encore [Stephan Krygsman & al. \(2004\)](#) indiquait que les usagers du transport en commun qui ont un permis de conduire valide sont plus susceptibles de se rendre aux stations de transport en commun que les usagers qui n'ont pas de permis. Les précédents résultats ne permettent pas de faire ce lien directement mais des études plus approfondies sur les données pourraient être envisagées pour tenter de l'établir.

Nombre de véhicule dans le foyer (HOUSEVEH) :

Quand on considère le nombre de véhicules dans un foyer, on s'attend à ce que la multimodalité diminue avec l'augmentation du nombre de véhicules ([Ralph Buehler & al., 2007](#)). Dans la littérature d'ailleurs, le nombre de véhicules disponibles dans le ménage à un impact négatif sur la multimodalité des individus ([Eva Heinen & Kiron Chatterjee, 2015](#)). Ces mêmes auteurs indiquent qu'un accès illimité à un véhicule domestique augmente considérablement la probabilité que tous les déplacements soient effectués à l'aide d'un véhicule et donc que la personne soit moins multimodale.

Tableau 30 : Résultats de la régression logistique générale et cotes relatives ("HOUSEVEH"), extrait du tableau 28

Catégorie	Estimate	Exp(estimate)	Effet sur les cotes (%)	p-value	Pertinence
0 véhicule (Référence locale)	0,00	-	-	-	-
1 véhicule	-1,32	0,27	-73,40	<0,01	***
2 véhicules	-1,94	0,14	-85,61	<0,01	***
3 véhicules	-1,92	0,15	-85,30	<0,01	***
4+ véhicules	-2,04	0,13	-86,93	<0,01	***

Les résultats montrent que le nombre de véhicules n'a pas vraiment d'influence sur la multimodalité des personnes, malgré une légère diminution de la multimodalité avec l'augmentation du nombre de véhicules. Cela reflète bien la mentalité des américains qui utilisent la voiture quoiqu'il arrive même lorsque le nombre de véhicules dans le foyer est faible. La culture de la voiture est tellement présente que beaucoup d'enfants, en âge d'avoir le permis (à 16 ans), ont leur propre voiture au lieu d'utiliser la voiture de leurs parents, ce qui est particulièrement le cas en dehors des villes selon [Marco Diana \(2012\)](#).

A l'inverse, comme le soulignait [Bai & Kattan \(2014\)](#), le partage d'une voiture (ex : empruntée par les enfants), favorise l'utilisation de transports en commun et donc la multimodalité des usagers. Cela qui est beaucoup plus le cas dans d'autres régions du monde comme en Europe ou en Asie où l'accès aux transports est plus facile et la voiture moins présente. Les foyers sans voiture sont en revanche bien plus multimodaux que les ménages possédant au moins une voiture, ce qui rejoint les conclusions de ([Ralph Buehler & al., 2007](#)). Les américains sans voiture sont donc contraints d'utiliser des alternatives, ce qui implique généralement l'utilisation de transports en communs. Ainsi, même si certains auteurs comme [Zihao An & al. \(2021\)](#) indiquaient une évolution positive de la multimodalité chez les automobilistes des pays développés, les américains restent très attachés à ce mode de transport, très ancré dans leur culture, en plus d'être un mode flexible et commode ([Sungyop Kim & al., 2007](#)).

Age (R_AGE) :

L'âge est une variable particulière dans l'étude, car elle est considérée comme *quasi continue* avec un nombre important de catégories (correspond aux valeurs uniques d'âge de la population interrogées de 5 à 92 ans). Par soucis de clarté, on considère cette variable comme une catégorie en tant que telle qui est ensuite comparée aux autres variables dans la régression logistique globale :

Tableau 31 : Résultats de la régression logistique générale et cotes relatives ("R_AGE"), extrait du tableau 28

Catégorie	Estimate	Exp(estimate)	Effet sur les cotes (%)	p-value	Pertinence
R_AGE	-0,02	0,98	-1,95	<0,01	***

La catégorie de référence ici est l'âge de 5 ans à partir de laquelle tous les autres âges sont comparés à la variable d'intérêt principale (la multimodalité). La variable liée aux catégories d'âge et liée à l'âge donnait déjà une certaine tendance :

Tableau 32 : Proportion de multimodaux par catégories d'âge

Variables	Multimodaux	% du total	Non-multimodaux	% du total	Total
(5,16] ans	124	9,67	1158	90,33	1282
(16,35] ans	81	3,93	1982	96,07	2063
(35,50[ans	34	1,60	2092	98,40	2126
(50,92] ans	57	1,11	5063	98,89	5120
Total	296		10295		10591

Une pré analyse, basée sur les catégories d'âge, permet déjà de voir que chez les jeunes (surtout avant l'âge du permis de conduire à 16 ans), près de 10% sont multimodaux, ce qui est 3 fois plus par exemple que les jeunes adultes (catégorie des (16,35] ans) relativement au nombre totale de personne dans chaque tranche d'âge.

Pour ce qui est de la variable *Age*, on constate que, la multimodalité diminue avec le temps (de 1,95% par année à partir de l'âge de 5 ans), ce qui est conforme aux résultats trouvés par [Sungyop Kim & al. \(2007\)](#) ou [Yang Jiang & al. \(2012\)](#) ou encore [Blumenberg & Pierce \(2013\)](#). A mesure qu'ils vieillissent, les américains sont en effet plus monomodaux de par leur travail ou à cause d'une santé qui se dégrade avec le vieillissement (limitant de fait l'utilisation des mobilités des mobilités douces ([Eva Heinen & Kiron Chatterjee, 2015](#))). Ceci expliquerait pourquoi leurs déplacements sont généralement plus courts d'après [Stephan Krygsman & al. \(2004\)](#).

Cependant, les résultats ne permettent pas d'étudier la multimodalité selon l'âge avec plus de détail (ex : par tranches d'âge). Des analyses complémentaires sont donc nécessaires pour vérifier que les voyageurs plus âgés voyagent davantage à des fins récréatives ([Eric J.E. Molin & al., 2010](#)) ou qu'ils sont plus multimodaux que les personnes ayant un emploi ([Thomas Klinger, 2017](#)).

Sexe (R_SEX) :

La concaténation des catégories de variables a conduit à retenir seulement trois catégories finales pour classer les personnes selon leur genre. En plus des catégories *Femme* et *Homme* classiques, on retrouve

une catégories *Autres* qui inclus notamment les personnes se définissant comme à la fois femme et homme ou aucun des deux.

On s'attendait à ce que les hommes voyagent en moyenne plus longtemps que les femmes (Stephan Krygsmann & al., 2004, Eric J.E. Molin & al., 2010). Mais dans le même temps, on s'attendait également à ce que les femmes combinent en général plus de modes de transport que les hommes, en particulier dans le cas des transports publics (Sungyop Kim & al., 2007, Ralph Buehler & al., 2007).

Tableau 33 : Résultats de la régression logistique générale et cotes relatives ("*R_SEX*"), extrait du tableau 28

Catégorie	Estimate	Exp(estimate)	Effet sur les cotes (%)	p-value	Pertinence
Female (Référence locale et globale)	0,00	-	-	-	-
Male	0,39	1,47	47,02	<0,01	**
Other	0,80	2,21	121,48	0,08	.

Etonnamment, les hommes sont en moyenne plus multimodaux que les femmes dans notre étude (+ 47%) et cela va à l'encontre de la littérature (Thomas Klinger, 2017, Eric J.E. Molin & al., 2010). On précise que les auteurs parlent surtout de situations hors des Etats-Unis (ex : en Europe) où les ménages possède moins de voiture et l'homme est en général prioritaire pour son utilisation sur la femme ce qui rend cette dernière de fait plus multimodale (Ralph Buehler & al., 2007). Ici on la vue, chaque membre du ménage possède généralement sa voiture et donc la femme n'est pas contrainte à utiliser d'autres modes comme les transports en communs par exemple. La société américaine étant très patriarcale, c'est en générale les femmes qui s'occupent des courses et bien des enfants et la voiture individuelle sera en général privilégiée sur les autres modes de transports dans ce cas.

D'autres auteurs, comme Sungyop Kim & al. (2007) et Ralph Buehler & al. (2007) indiquent que les hommes ont tendance à utiliser davantage la voiture que les femmes mais dans le but d'atteindre d'autres types de transport comme le vélo ou même les transports en communs par exemple. De même, on l'a vu, les hommes utilisent moins la voiture que les femmes mais plutôt le vélo ou le taxi pour se rendre sur leur lieu de loisir (Eric J.E. Molin & al., 2010). D'autres facteurs comme l'insécurité des femmes, notamment le soir, limite leur déplacement alors que les hommes ont moins ce problème, ce qui peut également expliquer les résultats. En revanche, on ne peut rien dire sur la multimodalité associée à la catégorie *Autres* car elle n'est pas significative (*p-value* trop faible). Ainsi, la multimodalité des personnes se définissant ni comme homme, ni comme femme, ou bien comme à la fois femmes et hommes, ne peut pas être évaluée dans ce modèle tel qu'il a été encodé.

Distance (*Distance Intervals*) :

Chaque trajet était associé à une distance totale définie comme la somme des miles parcourus par l'utilisateur durant une journée (somme de ses déplacements). Pour faciliter l'analyse et le traitement des données, la distance a été réduite à 4 catégories permettant de classer les usagers.

Tableau 34 : Résultats de la régression logistique générale et cotes relatives ("*Distance_Intervals*"), extrait du tableau 28

Catégorie	Estimate	Exp(estimate)	Effet sur les cotes (%)	p-value	Pertinence
(0,10[miles (Référence locale)	0,00	-	-	-	-
(10,25[miles	3,00	20,17	1917,47	<0,01	***

(25,50[miles	1,28	3,60	260,36	<0,01	***
(50,5000[miles	2,08	8,03	702,59	<0,01	***

Les résultats indiquent que la catégorie (10,25[miles a le plus d'influence sur la multimodalité des usagers. En d'autres termes, les personnes voyageant sur des distances moyennes sont plus susceptibles d'être multimodaux (presque 1,5 fois plus que les petits trajets inférieurs à 25 miles). Viennent ensuite ces mêmes petits trajets (inférieurs à 25 miles) pour lesquels les personnes sont aussi susceptibles d'être multimodaux.

En revanche, les trajets sur de longues distances (supérieurs à 25 miles) favorisent moins la multimodalité, en particulier pour les distances entre 25 et 50 miles (près de 3 fois moins que pour les trajets sur 10,25[miles). Ces résultats rejoignent ceux de [Ralph Buehler & al., 2007](#) qui indiquait notamment que la distance était en moyenne plus importante chez les automobilistes monomodaux que multimodaux. De même, selon [Blumenberg & Pierce \(2013\)](#), les déplacements sur de longues distances restent majoritairement associés à la voiture, voire l'avion aux Etats-Unis.

On remarque cependant que la multimodalité tend à réaugmenter lorsque la distance est très grande (> supérieurs à 50 miles), ce qui correspond par exemple à des trajets effectués entre plusieurs villes différentes mais situées dans une même zone géographique (ex : une aire urbaine, un County américain, etc.). Cela est notamment le cas pour les personnes faisant des allers-retours entre leur domicile et leur travail ou bien entre les centres-villes qui concentrent les activités et les villes en périphérie.

En revanche, on s'attendait à ce que la catégorie (0,10[miles soit celle qui favorise le plus la multimodalité et cela est principalement dû au fait que l'on n'a pas considéré la marche comme mode de transport propre. En effet, la marche est privilégiée dans le cas où les distances à parcourir sont très faibles ([Stephan Krygsman & al., 2004](#), [Eric J.E. Molin & al., 2010](#)), mais celle-ci implique un certain effort physique. Reste que l'obésité est un vrai sujet aux Etats-Unis et ce facteur nuit à l'utilisation des modes doux même au cœur des villes, là où la voiture sera privilégiée même pour de très courts trajets.

Jour de la semaine (TRAVDAY) :

A défaut d'avoir une date précise pour mettre en lumière la multimodalité des personnes sur une période de l'année, l'étude permet d'évaluer ce paramètre sur les jours de la semaine. La semaine (du lundi au vendredi) est associée à la période de travail et donc d'aller-retours entre le domicile et le lieu du travail notamment. Le samedi est généralement associé à certaines activités mais également aux courses et autres achats essentiels que l'on a pas le temps de faire durant la semaine de travail.

Tableau 35 : Résultats de la régression logistique générale et cotes relatives ("TRAVDAY"), extrait du tableau 28

Catégorie	Estimate	Exp(estimate)	Effet sur les cotes (%)	p-value	Pertinence
Sunday (Référence locale)	0,00	-	-	-	-
Monday	1,48	4,37	337,13	<0,01	***
Tuesday	1,19	3,27	227,48	0,01	**
Wednesday	1,56	4,78	377,56	<0,01	***
Thursday	1,24	3,45	245,47	<0,01	**
Friday	1,20	3,32	231,58	0,01	**
Saturday	1,44	4,22	321,75	<0,01	**

L'analyse des données tend à montrer que les usagers sont bien moins multimodaux le dimanche que les autres jours (samedi inclus), et cela rejoint la littérature (Ralph Buehler & al., 2007, Hye Kyung Lee & al., 2021). Ce jour est généralement associé à la détente à la maison ou aux voyages monomodaux (généralement en voiture), pour se rendre sur un lieu de loisir par exemple. Au contraire, les automobilistes sont plus multimodaux la semaine car ils combinent davantage la voiture avec la marche plutôt qu'avec d'autres modes de transports comme le bus ou le train (Ralph Buehler & al., 2007). De même, quand le nombre de déplacements quotidiens augmente, la personne est susceptible plus d'être multimodale (Blumenberg & Pierce, 2013). Les déplacements sont généralement plus importants la semaine plutôt que les weekends, ce qui tend à favoriser la multimodalité.

Quant à savoir quel jour les usagers sont plus multimodaux, d'autres tests spécifiques doivent être conduits pour répondre à cette question. Pour le moment, on remarque une propension à être légèrement plus multimodal le mercredi ou le lundi mais on ne peut pas affirmer que les gens sont drastiquement plus multimodaux ces jours-là que les autres. Une analyse plus poussée pourrait par exemple mettre en lumière les jours de télétravail par exemple qui tend à réduire drastiquement la multimodalité des personnes.

Temps de trajet (Travel time Intervals) :

La variable liée au temps de trajet total (TRVLCMIN) a été discrétisée en différents intervalles de temps pour faciliter le traitement sur R et l'analyse par la suite, en lien avec la littérature. Ce temps de trajet total a été calculé en sommant le temps entre les différents déplacements quotidiens mais sans prendre en compte le temps d'attente (ou de transfert) entre les différents modes. Intuitivement, on s'attend à ce que la multimodalité augmente avec l'augmentation du temps de transport (Riccardo Gallotti & Marc Barthelemy, 2014), car la probabilité de changer de mode de transport augmente.

Tableau 36 : Résultats de la régression logistique générale et cotes relatives ("Travel_time_Intervals"), extrait du tableau 28

Catégorie	Estimate	Exp(estimate)	Effet sur les cotes (%)	p-value	Pertinence
(0,30[minutes (Référence locale)	0,00	-	-	-	-
(30,60[minutes	0,09	1,09	9,28	0,60	
(60,120[minutes	-0,83	0,44	-56,22	<0,01	***
(120,1140[minutes	-1,28	0,28	-72,19	<0,01	***

La multimodalité est plus importante pour les trajets inférieurs à 30 minutes (temps de transport total), ce qui correspond à des trajets surtout urbains sur des distances généralement assez faibles. Un pic de multimodalité est indiqué pour les trajets entre 30 minutes et 1 heure mais ce résultat n'est pas assez significatif dans notre modèle (*p-value* trop faible) et ne sera donc pas considéré. Reste que les trajets très courts sont généralement monomodaux et ce quel que soit le mode (voiture, vélo, transports en commun, marche etc.) car les usagers ne veulent pas perdre de temps lors d'un changement de transport, en particulier lorsque la distance à parcourir est faible.

Au contraire, selon Riccardo Gallotti & Marc Barthelemy (2014), les trajets un peu plus longs, dans des zones où le réseau de transports est dense, offrent plus d'alternatives aux usagers en cas de problème sur leur ligne de transport, mais cette affirmation est impossible à vérifier ici comme on l'a vu. Au-delà de cette durée, les transports en communs ne sont pas envisagés mais plutôt la voiture ou même le taxi (Stephan Krygsman & al., 2004). Ces temps de transport très longs sont associés à des

déplacements sur de longues distances, généralement hors des réseaux de transport denses des centres-villes.

On rappelle que l'avion a été volontairement exclu de l'étude car on privilégie l'étude de la multimodalité dans les centres urbains. Or l'avion est l'un des modes les plus utilisés par les américains, même pour les trajets à l'intérieur d'un même état et ce mode est bien plus démocratisé que dans d'autres régions du monde (ex : en Europe). A l'inverse, les transports en commun hors centre-ville sont très peu développés, notamment les systèmes de train vers la banlieue ou les zones plus rurales (notamment les TER).

Ces solutions existent mais sont bien moins efficaces pour le transport de passagers que les solutions en Europe par exemple et sont donc associées à des temps de transport très long (parfois plusieurs heures). Même au cœur des villes, les services de transports en commun américains sont moins efficaces car ils ne disposent pas d'infrastructures dédiées, comme des voies réservées pour les bus autobus ou une meilleure gestion globale de la circulation en ville (Ehab I. Diab & Ahmed M. El-Geneidy, 2012).

Densité urbaine (URBANIZE) :

L'étude cherche principalement à étudier la multimodalité dans les centres urbains donc on a regroupé les personnes selon 3 catégories principales de densité auxquelles s'ajoute une catégorie *Autres densité (Other Density)* incluant notamment les zones rurales à faible densité. On précise que l'on ne parle pas de nombre d'habitants rapporté à une surface donnée mais bien à un nombre absolu de personnes vivant dans une ville donnée. La catégorie de référence (*Intercept*) correspond aux villes petites et moyennes par exemple. On s'attend à ce que la multimodalité augmente avec l'augmentation du nombre d'habitants (Ralph Buehler & al., 2007).

Tableau 37 : Résultats de la régression logistique générale et cotes relatives ("URBANIZE"), extrait du tableau 28

Catégorie	Estimate	Exp(estimate)	Effet sur les cotes (%)	p-value	Pertinence
50,000-499,000 (Référence locale)	0,00	-	-	-	-
500,000-999,999	-0,47	0,63	-37,26	0,16	
1,000,000 or more	0,46	1,59	58,75	0,01	*
Other Density	0,54	1,71	70,82	0,01	**

Les résultats montrent que la multimodalité est très importante dans les plus grands centres urbains (supérieur à 1 millions d'habitants) qui concentrent les activités et qui sont en général très denses (ex : New-York). Cela induit de fait un développement plus important de l'offre en transport en commun (ex : bus, métro), en particulier au cœur des villes où les rues sont plus étroites et la congestion du trafic limite l'utilisation de la voiture comme le soutenaient Ralph Buehler & al. (2007) ou encore Blumenberg & Pierce (2013). Le manque de place dans les grandes villes et la concentration des activités limitant les distances ce qui favorisent également l'utilisation de modes alternatifs, parmi lesquels, les services de micromobilité partagée (vélos, trottinettes électriques, e-scooters, etc.) qui contribuent positivement à la multimodalité (Xavier Bach & al., 2023).

L'augmentation de la densité favorise également la marche et donc la multimodalité car plus d'activités sont accessibles sur de faibles distances (Stephan Krygsman & al., 2004). La marche n'a pas été comptabilisée comme mode propre dans l'étude mais on comprend qu'elle constitue le moyen de

transport privilégié pour faire le lien entre d'autres modes (ex : de la voiture au train ou du vélo au bus). D'après les auteurs précédemment cités, la marche permet de réduire le temps de trajet lorsque la densité augmente, bien-sûr jusqu'à une certaine limite (congestion piétonne trop importante). Reste qu'une bonne planification des transports et des aménagements compacts favorisant les usages mixte en incitant les passagers à utiliser les modes doux au lieu de conduire ([Mode Choice to Modal Diversio & al., 2017](#)).

On remarque également que, comme on pouvait s'y attendre, la multimodalité diminue globalement avec la diminution du nombre d'habitant lorsque l'on reste dans des pôles urbains supérieurs à 50 000 habitants, ce qui fait écho aux résultats de [Stephan Krygsman & al. \(2004\)](#) ou [Marco Diana \(2012\)](#). Cela est dû au fait qu'en dehors des centres-villes américains majeurs, les villes sont en général très peu denses et l'étalement urbain y est très présent, ce qui favorise l'utilisation de véhicules individuels. Les Etats-Unis sont un très grand pays dans lequel de nombreux états sont structurés autour d'un ou deux grands pôles urbains qui concentrent les personnes et les activités, alors que le reste du territoire est bien moins densément peuplé (en particulier dans le centre du pays).

Cela s'explique également par des temps d'accès et de sortie aux transports plus importants dans le cas d'une densité urbaine moyenne ou bien celui d'une faible densité et, comme l'indiquaient [Stephan Krygsman & al. \(2004\)](#), cela nuit à l'utilisation d'alternatives à la voitures et donc à la multimodalité d'après [Stephan Krygsman & al. \(2004\)](#). Ces cas de densité sont aussi associées aux vastes zones pavillonnaires (population souvent aisée) en périphérie des grands centres urbains dans laquelle le comportement monomodal est privilégié par l'utilisation quasi exclusive de la voiture ([Blumenberg & Pierce, 2013](#)).

Le territoire est tellement vaste qu'il est difficile de développer un réseau de transports en commun aussi structurant qu'en Europe ou en Asie, ce qui explique d'ailleurs pourquoi l'avion est si développé aux Etats-Unis plus qu'ailleurs. On peut cependant relever une part plus importante de multimodaux chez les personnes vivant dans des centres urbains de moins de 50 000 habitants ou à la campagne (catégorie *Other*). Cette anomalie, par rapport à ce qui a été dit jusque-là, peut faire référence à toutes les personnes qui sont déposées à la gare la plus proche de chez eux pour aller ensuite travailler dans les grands centres urbains. En effet, les zones à faible densité encouragent la construction de TLR (train rapide) en prévision de l'augmentation de la population et de l'emploi à l'avenir ([Sungyop Kim & al., 2007](#)).

Cela concerne aussi les étudiants qui ne sont pas encore en âge de conduire et qui utilisent le vélo pour se rendre à l'arrêt de bus ou la gare la plus proche (généralement au cœur des petites villes ou villages). Cela peut également s'expliquer par la localisation de plus petites villes en périphérie directe de très grands centres urbains qui bénéficient de la proximité du réseau de transport de l'agglomération principale voire qu'y sont directement connectées (bus, métro, train, etc.). On notera cependant que l'on ne peut rien dire pour les villes entre 500,000 et 999,999 car le résultat n'est pas significatif.

But du voyage (WHYTO) :

Cette variable est sans doute celle qui a subi le plus d'étapes de concaténation des différentes catégories de base afin de pouvoir faire fonctionner le modèle. 3 catégories ont été conservées telles quelle car elles ont une influence très importante sur la multimodalité des personnes (*Regular activities at home, Attend school as a student* et *Change type of transportation*). Les deux autres rassemblent toutes les autres catégories de départ en les séparant les activités obligatoires (*Mandatory*) de celles

qui ne le sont pas (*Non mandatory*). Enfin, on note que la catégorie liée aux activités de la vie courante (courses, nettoyage, etc.) a été choisie arbitrairement par le logiciel R comme intercept.

Tableau 38 : Résultats de la régression logistique générale et cotes relatives ("WHYTO"), extrait du tableau 28

Catégorie	Estimate	Exp(estimate)	Effet sur les cotes (%)	p-value	Pertinence
Regular activities at home (Référence locale)	0,00	-	-	-	-
Attend school as a student	0,62	1,85	85,07	0,26	
Change type of transportation	1,95	7,05	605,24	<0,01	***
Other Mandatory	0,11	1,12	11,69	0,84	
Non mandatory	3,50	33,14	3214,41	<0,01	***

Comme on pouvait s'y attendre, le fait de changer de moyen de transport influence grandement la multimodalité des utilisateurs car c'est souvent pour passer de la voiture au transport en commun. Les personnes se rendant à l'école tendent à sembler en apparence être plus multimodales que les autres, car ces personnes ne possèdent généralement pas de voiture. En revanche, les résultats ne sont pas assez significatifs pour pouvoir conclure.

Sont incluses dans *Other Mandatory* ou (*Autres obligatoire*) les activités liées au travail notamment ou encore à la santé qui contribue également à la multimodalité des personnes mais dans une moindre mesure. Là aussi, les résultats ne sont pas assez significatifs pour être considérés même si on peut noter un niveau de multimodalité plus faible que la plupart des autres catégories. La littérature indiquait également que la multimodalité était relativement faible dans le cadre du travail car les usagers privilégient les trajets les plus directs possibles sans trop de changements pour éviter les retards (Sungyop Kim & al., 2007, Eric J.E. Molin & al., 2010).

D'un autre côté, les activités non essentielles (*Non mandatory*) telles les loisirs, les achats non essentiels ou les visites de proches ont un impact important sur la multimodalité des usagers. Nombre de ces trajets sont effectués en voiture selon Stephan Krygsman & al. (2004), mais contrairement à la question du travail, les usagers sont moins affectés dans ce cas par les retards sur les lignes de transports. Les modes les moins coûteux sont souvent privilégiés dans le cas d'activités récréatives, car, comme l'indiquaient Eric J.E. Molin & al. (2010) et Chieh-Hua Wen & al. (2012). Ainsi, les usagers préféreront mettre plus de temps en train ou en bus que de devoir payer l'essence d'un trajet en voiture plus rapide et plus cher, en particulier lorsque la contrainte de temps n'est pas un problème. Ce résultat fait aussi écho au travail de Zhang et al. (2014) qui indiquaient que le vélo est généralement assez présent dans le cadre d'activités non obligatoires.

Les activités en lien avec le foyer (tâches domestiques, courses, etc.) ont l'impact le plus faible sur la multimodalité des usagers. Là aussi les trajets sont quasi exclusivement effectués en voiture, même si le vélo ou la marche à pied peuvent être aussi impliqués et ce plutôt dans les centres urbains (Zhang et al., 2014). Ces mêmes auteurs indiquent que le vélo est plutôt perçu comme positif pour la santé ou encore la protection de l'environnement (véritable sujet de société) mais les lobbies de l'automobile ainsi que les habitudes de déplacements des américains tendent à limiter la démocratisation de ce mode de transport. De plus, les routes ne sont pas adaptées à la circulation cycliste aux Etats-Unis par rapport à d'autres régions du monde (ex : aux Pays-Bas) et ce même dans les très grandes villes.

4- CONCLUSION

La multimodalité constitue l'un des enjeux majeurs de demain notamment par son lien non négligeable avec la question environnementale. Une meilleure interconnexion entre les différents modes et un réseau de transport optimisé permettraient de réduire drastiquement la friction lors des trajets et ainsi encourager l'utilisation d'alternatives aux véhicules individuels. La multimodalité passe également par de la publicité et par des incitations auprès de la population afin que les usagers changent leur manière de se déplacer. La multimodalité reste cependant sous-développée, même si elle est encouragée de plus en plus ces dernières années, en particulier dans les espaces urbains et économiques denses.

Ce travail visait à faire un état des lieux de la multimodalité à l'échelle d'un territoire avec l'exemple des Etats-Unis, pays développés mais également bastion incontestable du voyage en voiture personnelle. Ce pays, pourtant première puissance économique mondiale, est bien moins étudié que d'autres pays dans le monde sur ce sujet, comme en Asie ou en Europe par exemple, où l'offre de transport est plus conséquente. Il a donc fallu d'une part évaluer la part de personnes multimodales dans la population américaine, ce qui implique de fixer des critères arbitraires permettant de définir cette multimodalité. D'autre part, ce travail visait à mettre en lumière les facteurs influençant le plus la multimodalité des usagers. Les données fournies par le NHTS donnent des informations très diverses, tant sur le déplacement (facteurs liés au transports et reconstitution du trajet) que sur les personnes interrogées (facteurs sociaux économiques). En revanche, les données ne permettent pas d'études plus approfondies sur la perception des transports sur les usagers ou encore le comportement même des personnes (facteurs psychologiques et culturels par exemple).

L'analyse des données a été conduite à l'aide de plusieurs logiciels, incluant un modèle de régression logistique, avec chacun leurs limites (notamment techniques). Les données de base ont dû être fortement traitées avant d'être injectées dans le modèle au vu du nombre important de variables et de catégories associées. Cela a permis de faciliter l'analyse par la suite même si cela implique une légère perte en précision sur les données. D'autres modèles plus puissants pourraient être utilisés pour mieux croiser les données entre elles, afin de créer de nouvelles variables d'intérêt, et ainsi compléter cette analyse (informations complémentaires sur les personnes comme sur les déplacements en eux-mêmes). Reste que le modèle utilisé est plutôt fiable et celui-ci offre une capacité prédictive satisfaisante permettant par la suite de faire le lien avec la littérature lors de la phase d'analyse.

La proportion de personnes multimodales est très faible par rapport à la taille de l'échantillon de base (moins de 3%), ce qui indique clairement que les Américains ne sont pas vraiment multimodaux. Dans le même temps, ce déséquilibre entre multimodaux et non multimodaux a eu une influence sur la pertinence du modèle et a conduit dans une certaine mesure à une mauvaise classification de certains usagers comme des personnes multimodales (et inversement). Cela n'a pas pour autant remis en cause la pertinence globale du modèle. On note également que ce résultat a été obtenu à partir de la définition de la multimodalité dans le modèle, et celle-ci est basée sur des hypothèses fortes qui diffèrent des autres définitions établies dans la littérature. Une redéfinition du caractère multimodal d'un trajet permettrait potentiellement d'augmenter le nombre de personnes multimodales (ex : en incluant la marche comme mode propre), mais la proportion de multimodaux américains restera toujours très faible quoi qu'il arrive (de l'ordre de quelques pourcents sur la population totale).

Ce résultat n'est pas aberrant dans la mesure où, malgré le développement récent des nouvelles formes de mobilité et les incitations pour l'utilisation des transports en communs, la voiture reste reine aux Etats-Unis. Au-delà du lobbying pour ce secteur, c'est un véritable mode de vie qui est très ancré dans la culture américaine, berceau du capitalisme, où la société s'est construite autour de la voiture et où

Le sentiment de liberté associé à la conduite est très présente dans les mœurs (et ce depuis la fin de la seconde guerre mondiale). La taille et la configuration du pays favorise plus l'avion pour les très longues distances et principalement la voiture pour les autres distances (moyennes et longues en particulier) et les modes alternatifs se retrouvent principalement dans les grands centres urbains qui concentrent les activités et les personnes. Face à ces constats, on remarque quand même une évolution positive de la multimodalité, notamment chez les jeunes, même si cette augmentation reste marginale ramenée à la population totale. La multimodalité s'étend également aux autres espaces urbains plus secondaires (en dehors des centres-villes) et est de plus en plus envisagée pour des longs trajets sur de grandes distances.

La population américaine, on l'a vu, est assez peu multimodale mais cette tendance peut être inversée, notamment par une sensibilisation accrue des pouvoirs publics auprès des usagers et par le développement d'un réseau de transports denses et optimisé. On a vu au travers de ce travail que la voiture dominait largement dans l'ensemble des modes transports et cela constitue un problème tant pour l'environnement que pour les usagers mêmes (ex : à cause du problème de congestion sur la route). Les réseaux de transport en commun (bus, métro, trains, tramways, etc.) doivent être davantage développés dans les villes (même moyennement denses) afin d'offrir une alternative crédible à la voiture. Ces moyens de transport principaux peuvent alors être couplés à des services de micromobilité, souvent électriques (vélo, trottinettes, etc.), pour effectuer les derniers miles dans les centres-villes où sont généralement concentrées les activités. Dans les zones moins denses, des parkings de délestage à proximité des gares, des services de navettes domicile-gare ou encore l'autopartage pourraient favoriser une utilisation plus multimodale de la voiture.

La commodité est primordial dans le choix du mode de transport et tout le système de transport doit être revu pour limiter au maximum la friction entre les différents modes, surtout face à la voiture qui est souvent privilégiée comme option moins stressante et qui requiert moins d'efforts physiques que les autres modes et ce, même si le trajet est légèrement plus long. A ce titre, des services d'abonnements donnant accès à tous les types de transport sur une même zone géographique peuvent être envisagés, ce qui nécessite cependant une harmonisation de l'ensemble des lignes et modes, souvent gérés par des compagnies différentes. L'incitation à l'utilisation d'alternatives à la voiture seule passe également par des prix compétitifs sur les titres de transport ou bien des aides financières pour l'achat de nouveaux moyens de locomotion (ex : micro mobilités). La notion de confort est d'autant plus importante chez les personnes âgées qui sont, comme on l'a vu, les personnes les moins multimodales. Un système d'information clair en temps réel et des mesures d'assistance dans les transports (ex : aide à la personne) pourraient rendre les transports en commun plus attractifs auprès de cette population. Enfin, le temps de transfert entre les modes doit être réduit au maximum sur l'ensemble du trajet pour que les usagers prennent conscience du gain réel de temps que procurent les trajets multimodaux.

Concernant l'analyse des données, celle-ci a conduit à ne réellement étudier qu'un nombre restreint de variables par rapport à l'ensemble des variables sélectionnées au départ (passage de 20 à 9), ou même par rapport à l'ensemble des indicateurs fournis par le NHTS. Cela est imputé au choix du modèle lors de l'analyse et également aux données en elles-mêmes. D'autres modèles plus avancés pourraient être développés afin de comparer plus de variables à la fois, montrant ainsi le rôle d'un nombre plus important de facteurs de tous types sur la multimodalité des personnes interrogées. On aurait pu également effectuer une comparaison binaire de catégorie à catégorie à l'intérieur d'une régression logistique plutôt que de confronter toutes les variables les unes par rapport aux autres. La collecte des données en elle-même pourrait aussi être revue afin d'obtenir un échantillon encore plus large de la population, ce qui améliorerait de fait le modèle et les prédictions statistiques associées.

Par ailleurs, ce travail s'est concentré sur l'analyse de certains facteurs directement disponibles ainsi que d'autres, calculés ou déduits à partir des données de base. L'accès à l'heure de départ et d'arrivée pour chaque déplacement aurait permis de distinguer plusieurs trajets distincts au sein d'une même journée, notamment si le temps d'attente est suffisamment long. Cela supposerait cependant de fixer une limite arbitraire de temps pour séparer les différents sous trajets au sein d'une même journée (ex : trajet le matin pour aller au travail puis le soir pour revenir à la maison). Des études complémentaires pourraient également être conduites pour mettre en lien le déplacement d'un usager considéré par rapport à celui des autres usagers avec qui la personne voyage, notamment celles faisant partie du même foyer. Ces données sont disponibles dans les données de base mais elles concernent surtout l'analyse des déplacements individuels alors que ce TFE s'est concentré sur l'analyse des trajets dans leur globalité (présence de personnes différentes potentiellement à chaque nouveau déplacement).

L'idée derrière la seconde phase de l'analyse était de retrouver, au sein des données courantes, des facteurs déjà présents dans la littérature, en particulier pour confirmer ou infirmer les précédents résultats. Cependant, d'autres facteurs, très peu abordés par la littérature, étaient également disponible dans les données de base, comme par exemple : le handicap, le Covid et son impact sur les déplacements, le paiement d'une partie du titre de transport par l'employeur ou encore le détail précis de l'utilisation du véhicule. Concernant la collecte des données, l'absence de dates précises pour chaque trajet (seulement le jour de la semaine et le mois) empêche de comparer la multimodalité sur différentes périodes de l'année (entre les périodes de travail et les vacances ou les jours fériés par exemple). Enfin, on l'a vu, la question de la marche comme mode propre est très ambiguë dans les données de base et davantage de trajets et déplacements auraient pu être multimodaux si ce mode avait été clairement considéré comme un moyen de transport dès le départ.

Le travail de traitement et d'analyse des données visait à retravailler et compléter les données de bases du NHTS dans le but de réaliser une régression logistique. Le programme développé sur R pourrait être optimisé et amélioré afin de faire tout le travail de pré-traitement des données directement sur le logiciel de façon automatique, sans travail au préalable sur Excel. Ainsi, n'importe quelle personne souhaitant poursuivre ce travail aurait simplement à télécharger les données brutes sur le site du NHTS et à lancer le programme. Ce programme pourrait également être complété par davantage d'indicateurs visuels (courbes, graphiques, tableaux récapitulatifs) tant sur la partie traitement des données que la partie analyse et régression. Cela permettrait notamment à des personnes non scientifiques de formation, comme des citoyens lambda ou des décideurs, d'analyser les données collectées sur le terrain sans pour autant comprendre toute la logique du programme dans ses moindres détails. On pourrait également rajouter d'autres modèles au programme en plus du modèle de régression logistique afin de pouvoir comparer les résultats.

Au-delà de tout ce qui a été précédemment dit, d'autres études pourraient être effectuées afin de connaître l'influence de certains facteurs peu documentés dans la littérature, comme les facteurs psychologiques par exemple. Les questionnaires et méthodes de collecte de données sont aussi à repenser dans ce cas, car les réponses sont, alors plus sujettes à interprétation et à la subjectivité. Cependant, cela permettrait d'avoir une meilleure appréhension de la perception des modes de transport par les usagers et ainsi de mieux suivre l'évolution des tendances et de comportements. Car en effet, l'amélioration des réseaux de transport est souvent la priorité des décideurs politiques dans le cadre du développement de la multimodalité. Je pense cependant, qu'un travail de sensibilisation de la population à la multimodalité et l'incitation à l'utilisation d'alternatives à la voiture sont des solutions tout aussi efficaces (peut-être même plus efficaces).

5- ANNEXES

5.1 Compléments

5.1.1 Figures

	TRIPID	TDCASEID	PERSONID	PER_ID	HOUSEID	LOOP_TRIP	MULTIMODAL	TDAYDATE	DATE
1	1	130020101	1	1300201	13002	2	NA	202202	Februa
2	2	130020102	1	1300201	13002	2	NA	202202	Februa
3	1	130020201	2	1300202	13002	2	NA	202202	Februa
4	2	130020202	2	1300202	13002	2	NA	202202	Februa
5	3	130020203	2	1300202	13002	2	NA	202202	Februa
6	1	130020401	4	1300204	13002	2	NA	202202	Februa
7	2	130020402	4	1300204	13002	2	NA	202202	Februa
8	1	130160101	1	1301601	13016	2	NA	202202	Februa
9	2	130160102	1	1301601	13016	2	NA	202202	Februa
10	3	130160103	1	1301601	13016	2	NA	202202	Februa
11	1	130160201	2	1301602	13016	2	NA	202202	Februa
12	2	130160202	2	1301602	13016	2	NA	202202	Februa
13	3	130160203	2	1301602	13016	2	NA	202202	Februa
14	1	130390101	1	1303901	13039	2	NA	202201	January
15	2	130390102	1	1303901	13039	2	NA	202201	January
16	1	130420101	1	1304201	13042	2	NA	202202	Februa
17	2	130420102	1	1304201	13042	2	NA	202202	Februa
18	1	130420201	2	1304202	13042	2	NA	202202	Februa
19	2	130420202	2	1304202	13042	2	NA	202202	Februa
20	1	130430101	1	1304301	13043	1	NA	202202	Februa
21	1	130440101	1	1304401	13044	2	NA	202202	Februa

Figure 2 : Aperçu du tableau de référence "data_TRIP" sur l'interface de RStudio

5.1.2 Tableaux

Tableau 39 : Données non conservées (Persons_pub)

Paramètre	Description	Justification
BIKETRANSIT	Jours des 30 derniers jours de vélo utilisé	Pas de lien direct avec le trajet (sur la journée)
BIKESHARE22	Jours des 30 derniers jours du système de vélo partagé	Pas de lien direct avec le trajet (sur la journée)
CONDNIGH	Conduite limitée à la journée en raison d'une condition ou d'un handicap	Paramètre non étudié
CONDRIDE	Trajet abandonné du fait d'un handicap	Paramètre non étudié
CONDNONE	Voyage non affecté par une condition ou un handicap	Paramètre non étudié

CONDPUB	Utilisation moins fréquente de l'autobus ou du métro en raison d'une condition ou d'un handicap	Paramètre non étudié
CONDRF	Je préfère ne pas répondre si le voyage est affecté par une condition ou un handicap	Paramètre non étudié
CONDSHARE	Covoiturage utilisé en raison d'une condition ou d'un handicap	Paramètre non étudié
CONDSPEC	Utilisation de services de transport spéciaux en raison d'une condition ou d'un handicap	Paramètre non étudié
CONDTRAV	Réduction des déplacements en raison d'une condition ou d'un handicap	Paramètre non étudié
COV1_WK	Impact du Covid sur les trajets liés au lieu de travail	Paramètre non étudié
COV1_SCH	Impact du Covid sur les trajets liés à l'école/l'emplacement de la classe physique	Paramètre non étudié
COV1_PT	Impact de la COVID sur l'utilisation du transport en commun	Paramètre non étudié
COV1_OHD	Impact du COVID sur les achats en ligne pour la livraison à domicile	Paramètre non étudié
COV2_WK	Changements temporaires ou permanents sur les déplacements professionnels	Paramètre non étudié
COV2_SCH	Trajet pour aller à l'école, changement temporaire ou permanent	Paramètre non étudié
COV2_PT	Changements temporaires ou permanents à l'utilisation du transport en commun	Paramètre non étudié
DELIVER	Nombre de livraisons d'achats en ligne au cours des 30 derniers jours	Pas de lien direct avec le trajet (sur la journée)
DELIV_GOOD	Nombre de marchandises livrées au cours des 30 derniers jours	Pas de lien direct avec le trajet (sur la journée)
DELIV_FOOD	Nombre de fois où de la nourriture a été livrée au cours des 30 derniers jours	Pas de lien direct avec le trajet (sur la journée)
DELIV_GROC	Nombre de livraisons d'épicerie au cours des 30 derniers jours	Pas de lien direct avec le trajet (sur la journée)
DELIV_PERS	Nombre de fois où les services ont été fournis au cours des 30 derniers jours	Pas de lien direct avec le trajet (sur la journée)
DRIVINGOCCUPATION	Conduire pour le travail	Redondance avec <i>WHYTO</i>
DRIVINGVEHICLE	Véhicule conduit pour le travail	Redondance avec <i>VEHCASEID</i>
EMPLOYMENT2	Heures travaillées contre rémunération chaque semaine	Pas de lien direct avec le trajet (sur la journée)
EMPPASS	L'employeur paie pour un laissez-passer de transport en commun à prix réduit	Paramètre non étudié

ESCOOTERUSED	Nombre de jours pour l'utilisation de l'e-scooter dans les 30 derniers jours	Pas de lien direct avec le trajet (sur la journée)
GCDWORK	Distance orthodromique (miles) entre le domicile et le travail	Distance et statut professionnel déjà considéré
LAST30_TAXI	Service de taxi utilisé dans les 30 derniers jours	Pas de lien direct avec le trajet (sur la journée)
LAST30_RDSHR	Service d'auto partagé utilisé dans les 30 derniers jours	Pas de lien direct avec le trajet (sur la journée)
LAST30_ESCT	E-scooters utilisé dans les 30 derniers jours	Pas de lien direct avec le trajet (sur la journée)
LAST30_PT	Transport en commun utilisé le au cours des 30 derniers jours	Pas de lien direct avec le trajet (sur la journée)
LAST30_MTRC	Moto d'occasion au cours des 30 derniers jours	Pas de lien direct avec le trajet (sur la journée)
LAST30_WALK	Marche d'un endroit à l'autre au cours des 30 derniers jours	Pas de lien direct avec le trajet (sur la journée)
LAST30_BIKE	Vélo usagé au cours des 30 derniers jours	Pas de lien direct avec le trajet (sur la journée)
LAST30_BIKE	Vélos en libre-service d'occasion au cours des 30 derniers jours	Pas de lien direct avec le trajet (sur la journée)
LAST30_TAXI	Service de taxi utilisé au cours des 30 derniers jours	Pas de lien direct avec le trajet (sur la journée)
MCTRANSIT	Jours dans les 30 derniers jours moto utilisée	Pas de lien direct avec le trajet (sur la journée)
MEDCOND	Condition ou handicap qui rend le voyage difficile	Paramètre non étudié
MEDCOND6	Etat de santé du répondant	Paramètre non étudié
OUTOFTWN	Loin de chez soi toute la journée de voyage	Heures de départ et d'arrivée de chaque trajet déjà présentes
PARKHOME	Payer le stationnement à domicile	Paramètre non étudié
PARKHOMEAMT	Si le répondant paie pour se garer à la maison	Paramètre non étudié
PARKHOMEAMT_PAMOUNT	Coût du stationnement à la maison	Paramètre non étudié
PARKHOMEAMT_PAYTYPE	Durée du paiement	Paramètre non étudié
PAYPROF	A travaillé pour un salaire la semaine dernière	Paramètre non étudié
PTUSED	Jours des 30 derniers jours d'utilisation du transport en commun	Pas de lien direct avec le trajet (sur la journée)
QACSLAN1	Langue autre que l'anglais parlée à la maison	Paramètre non étudié
QACSLAN3	À quel point cette personne parle bien anglais	Paramètre non étudié
RET_HOME	Nombre de fois où l'achat en ligne a été retourné par ramassage à domicile	Paramètre non étudié
RET_PUF	Nombre de fois où l'achat en ligne a été retourné au bureau de poste/UPS/Fed Ex/ semblable	Paramètre non étudié

RET_AMZ	Nombre de fois les achats en ligne retournés au centre de dépôt Amazon	Paramètre non étudié
RET_STORE	Nombre de fois qu'un article en ligne a été retourné à un magasin	Paramètre non étudié
RIDESHARE22	Jours dans les 30 derniers jours de covoiturage utilisé	Pas de lien direct avec le trajet (sur la journée)
R_RACE_IMP	Origine ethnique (imputée)	Redondance avec <i>R_RACE</i>
R_RELAT	Relation avec le premier répondant	Pas de liens établis entre les participants dans l'étude
SAMEPLC	Raison de ne pas faire de voyages le jour du voyage	Paramètre non étudié
SCHOOL1	Inscrit à l'école ou à un programme académique	Redondance avec "EDUC"
STUDE	Description de l'école ou du programme scolaire	Paramètre non étudié
SCHTYP	Type d'écoles "K-12"	Paramètre non étudié
SCHOOL1C	Type d'écoles non "K-12"	Paramètre non étudié
SCHTRN1	Transport habituel vers l'école	Pas de lien direct avec le trajet (sur la journée)
TAXISERVICE	Jours des 30 derniers jours service de taxi utilisé	Pas de lien direct avec le trajet (sur la journée)
TRNPASS	Laissez-passer de transport en commun à prix réduit utilisé au cours des 30 derniers jours	Pas de lien direct avec le trajet (sur la journée)
USAGE1	Moins de voyages au cours des 30 derniers jours	Pas de lien direct avec le trajet (sur la journée)
USAGE2_1	Raison : moins de déplacements - plus de livraisons	Pas de lien direct avec le trajet (sur la journée)
USAGE2_2	Raison : moins de trajets – pas en sécurité	Pas de lien direct avec le trajet (sur la journée)
USAGE2_3	Raison : moins de voyages - ne se sentait pas propre/en bonne santé	Pas de lien direct avec le trajet (sur la journée)
USAGE2_4	Raison : moins de trajets - pas fiable	Pas de lien direct avec le trajet (sur la journée)
USAGE2_5	Raison : moins de voyages - je ne suis pas allé là où c'était nécessaire	Pas de lien direct avec le trajet (sur la journée)
USAGE2_6	Raison : moins de voyages - inabordable	Pas de lien direct avec le trajet (sur la journée)
USAGE2_7	Raison : moins de déplacements - problèmes de santé	Pas de lien direct avec le trajet (sur la journée)
USAGE2_8	Raison : moins de déplacements - pas de temps	Pas de lien direct avec le trajet (sur la journée)
USAGE2_9	Raison : moins de déplacements - autre	Pas de lien direct avec le trajet (sur la journée)
USAGE2_10	Raison : moins de voyages - COVID 19	Pas de lien direct avec le trajet (sur la journée)
USEPUBTR	Utilisation du transport en commun à la date du voyage	Redondance avec <i>PUBTRANS</i>
WALKTRANSIT	Jours dans les 30 derniers jours de marche utilisés	Pas de lien direct avec le trajet (sur la journée)

W_CANE	Utilise une canne ou un bâton de marche	Paramètre non étudié
W_CHAIR	Utilise un scooter manuel ou un fauteuil roulant	Paramètre non étudié
WHOPROXY	Qui répond au sondage	Seul les données renseignées sont intéressantes
WKFMHM22	Jours par semaine travaillés à domicile	Pas de lien direct avec le trajet (sur la journée)
W_NONE	N'utilise aucun appareil médical pour la mobilité	Paramètre non étudié
WRKLOC	Description du lieu de travail	Paramètre non étudié
WRKTRANS	Transport habituel pour se rendre au travail	Pas de lien direct avec le trajet (sur la journée)
W_SCCH	Utilise un scooter motorisé ou un fauteuil roulant	Paramètre non étudié
WTPERFIN	Poids sur 7 jours affecté à la personne	Chaque usager est regardé indépendamment
WTPERFIN5D	Poids sur 5 jours affecté à la personne	Chaque usager est regardé indépendamment
WTPERFIN2D	Poids sur 2 jours affecté à la personne	Chaque usager est regardé indépendamment
W_WKCR	Utilise un déambulateur ou des béquilles	Paramètre non étudié
W_VISIMP	Utilise des appareils pour aider les aveugles ou les malvoyants	Paramètre non étudié

Tableau 40 : Données non conservées (Households_pub)

Variables	Description	Justification
CNTTDHH	Nombre de déplacements des ménages le jour du voyage	On regarde les déplacement par personne et non par ménage
FLAG100	Tous les membres du foyer ont répondu au sondage ?	Paramètre non étudié
HHFAMINC_IMP	Revenu du ménage (imputé)	Redondance avec <i>HHFAMINC</i>
HHRELATD	Indicateur indiquant qu'au moins 2 personnes dans le foyer (household) sont apparentées	Composition familiale du ménage déjà présente
PPT517	Nombre de membres du ménage âgés de 5 à 17 ans	Composition familiale du ménage déjà présente
RESP_CNT	Nombre de personnes ayant répondu dans le ménage	Composition familiale du ménage déjà présente
WTHHFIN2D	Poids du ménage NAT sur 2 jours	Chaque usager est regardé indépendamment
WTHHFIN5D	Poids du ménage NAT sur 5 jours	Chaque usager est regardé indépendamment
YOUNGCHILD	Nombre de membres du ménage âgés de moins de 5 ans	Composition familiale du ménage déjà présente

Tableau 41 : Données non conservées (Vehicles_pub)

Variables	Description	Justification
ANNMILES	Estimation autodéclarée du nombre de milles annualisé	Paramètre non étudié
COMMERCIALFREQ	Au cours des 30 derniers jours, combien de jours le véhicule a-t-il été utilisé à des fins professionnelles ?	Paramètre non étudié
HHVEHCNT	Nombre total de véhicules dans le ménage	Paramètre non étudié
HHVEHUSETIME	Au cours des 30 derniers jours, combien de jours le véhicule a-t-il été utilisé pour d'autres activités ?	Paramètre non étudié
HHVEHUSETIME	Au cours des 30 derniers jours, combien de jours le véhicule a-t-il été utilisé pour les livraisons ?	Paramètre non étudié
HYBRID	Véhicule hybride	Paramètre non étudié
MAKE	ID du véhicule	Paramètre non étudié
VEHOWNED	Véhicule possédé depuis 1 an ou plus	Paramètre non étudié
VEHCOMMERCIAL	Véhicule utilisé à des fins professionnelles	Paramètre non étudié
VEHCOM_DEL	Véhicule utilisé pour la livraison	Paramètre non étudié
VEHCOM_OTH	Véhicule utilisé pour d'autres activités	Paramètre non étudié
VEHCOM_RS	Véhicule utilisé pour le covoiturage	Paramètre non étudié
VEHFUEL	Type de carburant	Paramètre non étudié
VEHAGE	Âge du véhicule, selon l'année-modèle	Paramètre non étudié
VEHOWNMO	Véhicules possédés depuis moins de 1 an - mois possédés	Paramètre non étudié
VEHYEAR	Année du véhicule	Paramètre non étudié
WHOMAIN	Conducteur principal du véhicule	Paramètre non étudié
WTHHFIN	Poids du ménage Natl sur 7 jours	Paramètre non étudié
WTHHFIN5D	Poids du ménage Natl sur 5 jours	Paramètre non étudié
WTHHFIN2D	Poids du ménage Natl sur 2 jours	Paramètre non étudié

Ajustement des variables :

Tableau 42 : Variables "NUMADLT/ADULTS" ajustée

Catégories	Catégories concaténées	Multimodaux	%	Non-Multimodaux	%	Total
1	-	62	3,28	1826	96,72	1888
2	-	183	2,68	6634	97,32	6817
3	-	43	3,30	1260	96,70	1303
4+	De 4 à 8	8	1,37	575	98,63	583
Total		296		10295		10591

Avec en gras les nouvelles catégories créées (concaténation des anciennes).

Tableau 43 : Variables "Age_Intervals" ajustée

Catégories	Catégories concaténées	Multimodaux	%	Non-Multimodaux	%	Total
(5,16[(5,13[; (13,16[124	9,67	1158	90,33	1282
(16,35[(16,25[; (25,35[81	3,93	1982	96,07	2063
(35,50[-	34	1,60	2092	98,40	2126
(50,65[-	33	1,28	2536	98,72	2569
(65,92]	(50,100[; (100,5000[24	0,94	2527	99,06	2551
Total		296		10295		10591

Avec en gras les nouvelles catégories créées (concaténation des anciennes).

Tableau 44 : Variables "Distance_Intervals" ajustée

Catégories	Catégories concaténées	Multimodaux	%	Non-Multimodaux	%	Total
(0,5[(0,0.5[; (0.5,2[; (2,5[40	1,76	2228	98,24	2268
(5,10[-	66	3,92	1617	96,08	1683
(10,25[-	113	3,76	2890	96,24	3003
(25,50[-	39	2,00	1909	98,00	1948
(50,5000[(50,100[; (100,5000[38	1,41	2651	98,59	2689
Total		296		10295		10591

Avec en gras les nouvelles catégories créées (concaténation des anciennes).

Tableau 45 : Variables "DRIVERS/DRVRCNT" ajustée

Catégories	Catégories concaténées	Multimodaux	%	Non-Multimodaux	%	Total
1	-	25	16,23	129	83,77	154
2	-	58	2,60	2177	97,40	2235
3	-	158	2,49	6197	97,51	6355
4+	De 4 à 11	43	3,37	1232	96,63	1275
Total		296		10295		10591

Avec en gras les nouvelles catégories créées (concaténation des anciennes).

Tableau 46 : Variables "EDUC/EDUCATION" ajustée

Catégories	Catégories concaténées	Multimodaux	%	Non-Multimodaux	%	Total
Less than high school	-	19	7,82	224	92,18	243

High school	Some high school, no diploma/GED ; High school graduate ; have diploma/GED	61	3,67	1599	96,33	1660
College	High school graduate, have diploma/GED ; Associates degree (2-year)/ trade school certificate	28	1,06	2607	98,94	2635
Bachelor's degree	-	60	2,13	2754	97,87	2814
Master's degree	-	28	1,77	1554	98,23	1582
Professional/Doctorate degree	-	6	0,96	620	99,04	626
Other Education	Devient <i>Valid skip</i>	94	9,12	937	90,88	1031
Total		296		10295		10591

Avec en gras les nouvelles catégories créées (concaténation des anciennes).

Tableau 47 : Variables "HHFAMINC/HH_INCOME" ajustée

Catégories	Catégories concaténées	Multimodaux	%	Non-Multimodaux	%	Total
Less than \$24,999	Less than \$10,000 ; \$10,000 to \$14,999 ; \$15,000 to \$24,999	40	5,35	707	94,65	747
\$25,000 to \$100,000	\$25,000 to \$34,999 ; \$35,000 to \$49,999 ; \$50,000 to \$74,999 ; \$75,000 to \$99,999	112	2,32	4721	97,68	4833
\$100,000 or more	\$100,000 to \$124,999 ; \$125,000 to \$149,999 ; \$150,000 to \$199,999 ; \$200,000 or more	143	2,93	4737	97,07	4880
Other income		1	0,76	130	99,24	131
Total		296		10295		10591

Avec en gras les nouvelles catégories créées (concaténation des anciennes).

Tableau 48 : Variables "HOUSEVEH/HHVEHCNT" ajustée

Catégories	Catégories concaténées	Multimodaux	%	Non-Multimodaux	%	Total
1	-	47	14,78	271	85,22	318
2	-	74	3,02	2373	96,98	2447
3	-	107	2,17	4829	97,83	4936
4+	De 4 à 7	46	2,51	1785	97,49	1831
Total		296		10295		10591

Avec en gras les nouvelles catégories créées (concaténation des anciennes).

Tableau 49 : Variables "PRMACT/ACTIVITY" ajustée

Catégories	Catégories concaténées	Multimodaux	%	Non-Multimodaux	%	Total
Unemployed	temporarily absent/unemployed	8	3,94	195	96,06	203
A homemaker	-	1	0,25	401	99,75	402
Going to school	-	26	9,39	251	90,61	277
Retired	-	23	1,02	2223	98,98	2246
Worker	Ajout de la catégorie <i>Employment</i>	106	1,86	5606	98,14	5712
Other Activity	-	132	7,54	1619	92,46	1751
Total		296		10295		10591

Avec en gras les nouvelles catégories créées (concaténation des anciennes).

Tableau 50 : Variables "LIF_CYL/FAMILY" ajustée

Catégories	Catégories concaténées	Multimodaux	%	Non-Multimodaux	%	Total
one adult, no children	one adult, no children ; one adult, retired, no children	40	13,51	1464	14,22	1504
2+ adults, no children	2+ adults, no children ; 2+ adults, retired, no children	71	23,99	4737	46,01	4808
one adult, youngest child 0-15	one adult, youngest child 0-5 ; one adult, youngest child 6-15	20	6,76	352	3,42	372
2+ adults, youngest child 0-15	2+ adults, youngest child 0-5 ; 2+ adults, youngest child 6-15	138	46,62	3113	30,24	3251
one adult, youngest child 16-21	-	5	1,69	111	1,08	116
2+ adults, youngest child 16-21	-	22	7,43	518	5,03	540
Other Family	-	0	0,00	0	0,00	0
Total		296		10295		10591

Avec en gras les nouvelles catégories créées (concaténation des anciennes).

Tableau 51 : Variables "R_AGE" ajustée

Catégories	Catégories concaténées	Multimodaux	%	Non-Multimodaux	%	Total
(5,16[(5,13[; (13,16[124	9,67	1158	90,33	1282
(16,35[(16,25[; (25,35[81	3,93	1982	96,07	2063
(35,50[-	34	1,60	2092	98,40	2126
(50,65[-	33	1,28	2536	98,72	2569
(65,92]	-	24	0,94	2527	99,06	2551
(5,16[-	124	9,67	1158	90,33	1282
Total		296		10295		10591

Avec en gras les nouvelles catégories créées (concaténation des anciennes).

Tableau 52 : Variables "R_RACE/RACE" ajustée

Catégories	Catégories concaténées	Multimodaux	%	Non-Multimodaux	%	Total
Asian	-	18	3,30	528	96,70	546
Black or African American	-	18	3,09	565	96,91	583
Hispanic	Ajout de la catégorie Hispanic	39	4,13	906	95,87	945
White	-	215	2,57	8151	97,43	8366
Other race	American Indian/Alaska Native ; Multiple race selected ; Native Hawaiian/Pacific Islander ; Other race	6	3,97	145	96,03	151
Total		296		10295		10591

Avec en gras les nouvelles catégories créées (concaténation des anciennes).

Tableau 53 : Variables "R_SEX/SEXE" ajustée

Catégories	Catégories concaténées	Multimodaux	%	Non-Multimodaux	%	Total
Female	-	124	2,38	5085	97,62	5209
Male	-	165	3,12	5116	96,88	5281
Other Sex	Refuse ; don't know	7	6,93	94	93,07	101
Total		296		10295		10591

Avec en gras les nouvelles catégories créées (concaténation des anciennes).

Tableau 54 : Variables "Travel_time_Intervals" ajustée

Catégories	Catégories concaténées	Multimodaux	%	Non-Multimodaux	%	Total
(0,15[(0,5[; (5,10[; (10,15[5	0,53	930	99,47	935
(15,30[(15,20[; (20,30[26	1,11	2322	98,89	2348
(30,45[-	45	2,84	1538	97,16	1583
(45,60[-	48	2,76	1692	97,24	1740
(60,120[-	111	4,08	2612	95,92	2723
(120,1140[(120,240[; (240,360[(360,11]	61	4,83	1201	95,17	1262
Total		296		10295		10591

Avec en gras les nouvelles catégories créées (concaténation des anciennes).

Tableau 55 : Variables "URBANIZE/DENSITY" ajustée

Catégories	Catégories concaténées	Multimodaux	%	Non-Multimodaux	%	Total
50,000-499,000	50,000-199,999 ; 200,000-499,999	48	2,10	2243	97,90	2291
500,000-999,999	-	13	1,28	1000	98,72	1013

1,000,000 or more	1,000,000 or more with heavy rail ; 1,000,000 or more without heavy rail	149	3,45	4173	96,55	4322
Other Density	A la place de <i>Not in urbanized area</i>	86	2,90	2879	97,10	2965
Total		296		10295		10591

Avec en gras les nouvelles catégories créées (concaténation des anciennes).

Tableau 56 : Variables "WHYTO/PURPOSE " ajustée

Catégories	Catégories concaténées	Multimodaux	%	Non-Multimodaux	%	Total
Regular activities at home	-	4	1,01	391	391	395
Work	Work from home (paid) ; Work at a non-home location ; Work activity to drop-off/pickup someone/something ; Volunteer activities (not paid)	49	1,93	2494	2494	2543
Attend school as a student	-	123	14,07	751	751	874
Health	Attend childcare or adult care ; Health care visit	12	2,40	489	489	501
Change type of transportation	-	25	32,89	51	51	76
Drop off/pick up someone (personal)	-	19	2,73	677	677	696
Spending	Buy meals ; Shop/buy/pick-up or return goods ; Other family/personal errands	23	0,74	3086	3086	3109
Recreational	Religious or other ; Rest or relaxation/vacation ; Exercise ; Recreational activities	32	1,76	1787	1787	1819
Visit friends or relatives	-	9	1,62	545	545	554
Other Purpose	A la place de "Other Purpose"	0	0,00	24	24	24
Total		296		10295		10591

Avec en gras les nouvelles catégories créées (concaténation des anciennes).

Ajustement complémentaires propres à la régression logistique :

Tableau 57 : Variables "Age_Intervals" ajustée

Catégories	Catégories concaténées	Multimodaux	%	Non-Multimodaux	%	Total
(5,16[-	124	9,67	1158	90,33	1282
(16,35[-	81	3,93	1982	96,07	2063
(35,50[-	34	1,60	2092	98,40	2126
(50,92]	(50,65[; (65,92]	57	1,11	5063	98,89	5120
Total		296		10295		10591

Avec en gras les nouvelles catégories créées (concaténation des anciennes).

Tableau 58 : Variables "Distance_Intervals" ajustée

Catégories	Catégories concaténées	Multimodaux	%	Non-Multimodaux	%	Total
(0,10[(0,5[; (5,10[106	2,68	3845	97,32	3951
(10,25[-	113	3,76	2890	96,24	3003
(25,50[-	39	2,00	1909	98,00	1948
(50,5000[-	38	1,41	2651	98,59	2689
Total		296		10295		10591

Avec en gras les nouvelles catégories créées (concaténation des anciennes).

Tableau 59 : Variables "Travel_time_Intervals" ajustée

Catégories	Catégories concaténées	Multimodaux	%	Non-Multimodaux	%	Total
(0,30[(0,15[; (15,30[31	0,94	3252	99,06	3283
(30,60[(30,45[; (45,60[93	2,80	3230	97,20	3323
(60,120[-	111	4,08	2612	95,92	2723
(120,1140[-	61	4,83	1201	95,17	1262
Total		296		10295		10591

Avec en gras les nouvelles catégories créées (concaténation des anciennes).

Tableau 60 : Variables "WHYTO/PURPOSE" ajustée

Catégories	Catégories concaténées	Multimodaux	%	Non-Multimodaux	%	Total
Regular activities at home	-	4	1,01	391	98,99	395
Other Mandatory	Work ; Health ; Drop off/pick up someone (personal)	80	2,14	3660	97,86	3740
Attend school as a student	-	123	14,07	751	85,93	874
Non mandatory	Spending ; Recreational ; Visit friends or relatives ; Other Purpose	64	1,16	5442	98,84	5506
Change type of transportation	-	25	32,89	51	67,11	76
Total		296		10295		10591

Avec en gras les nouvelles catégories créées (concaténation des anciennes).

5.1.3 Programmation sur R (code)

```
## >>> Libraries to install :
install.packages("readxl") # to read Excels files
install.packages("knitr") # for better table formalism
install.packages("kableExtra")
install.packages("dplyr")
install.packages("ISwR")
install.packages("tidyverse")
install.packages("caret")
install.packages("vcd")
install.packages("car")
install.packages("aod")
install.packages("gridExtra")
install.packages("hms")
install.packages("pROC")
install.packages("openxlsx")
installed.packages("pscl")
installed.packages("ResourceSelection")
install.packages("InformationValue")

library(readxl) # Charging library and then importing Excel file
data_TRIP <- read_excel("C:/Users/pierr/Downloads/TFE/Analyse données/Données
NHTS (US)/Données travaillées (EXCEL)/Fichiers Excel/Trip_pub
(retravaillé).xlsx",sheet="Trip_pub")

data_TRIP[data_TRIP == "-1"] <- NA # correcting negative values/missing values
data_TRIP[data_TRIP == "-7"] <- NA
data_TRIP[data_TRIP == "-8"] <- NA
data_TRIP[data_TRIP == "-9"] <- NA
data_TRIP[data_TRIP == "Valid skip"] <- NA

## >> Multimodality on the whole day ("data_TRIP_Person") :
library(ISwR) # Charging library
occurrence_1 = 0 # initial value to 0 (used to count total number of persons)
compiled_TRIPID_values <- list() # list of Trip IDs (empty)
compiled_PER_ID_values <- list() # list of Person IDs (empty)
compiled_TR_TYPE_values <- list() # list of type of transports
for (i in 1:nrow(data_TRIP)-1) { # to read "data_TRIP" row by row from 1 to
31073
  if (data_TRIP$PER_ID[i] != data_TRIP$PER_ID[i+1]) { # if it's a different
person...
    occurrence_1 <- occurrence_1 + 1 # then add 1 to the total number of
persons already considered
```

```

    compiled_TRIPID_values <-
append(compiled_TRIPID_values,data_TRIP$TRIPID[i]) # compilation of all Trip
IDs for a unique person
    compiled_PER_ID_values <-
append(compiled_PER_ID_values,data_TRIP$PER_ID[i]) # compilation of all
Person IDs for a unique person
    compiled_TR_TYPE_values <-
append(compiled_TR_TYPE_values,data_TRIP$TR_TYPE[i]) # compilation of all
type of transports for a unique person
  }
  library(dplyr) # Charging library
  library(hms) # Charging library
  SEQUENCE_TR <- data_TRIP %>% # table with the sequence of transports by
person
  group_by(PER_ID) %>% # uses Person ID as a reference to build the table
(first column)
  summarise(
    Trip_Sequence = paste(TRIPID, collapse = ", "), # list of trips
numbers for each Person
    Vehicle_Sequence = paste(TR_TYPE, collapse = ", "), # list of all
modes used for each Person
    VEHCASEID = paste(unique(VEHCASEID), collapse = ", "), # list of all
vehicles used for each Person (by vehicles IDs)
    PUBTRANS = ifelse(all(is.na(PUBTRANS)), NA,
paste(unique(na.omit(PUBTRANS)), collapse = ", ")), # avoid multiple time the
same value for "Public transport" variable
    TRPMILES = sum(TRPMILES), # total distance on the whole trip by person
    STRTTIME = STRTTIME[1], # first starting time of the whole trip by
person
    ENDTIME = tail(ENDTIME, 1), # last ending time of the whole trip by
person
    DWELTIME = sum(DWELTIME, na.rm = TRUE), # total dwell time on the whole
trip by person
    TRVLCMIN = sum(TRVLCMIN, na.rm = TRUE), ) %>% # total transport time on
the whole trip by person
    arrange(match(PER_ID, data_TRIP$PER_ID)) # follow the same order as in
data_TRIP with Person ID as a reference
  }
Tbl_Multi_Transport <- data.frame() # table of the persons that have at least
2 types of transports (multimodal or not)
Tbl_Multi_Transport <- data.frame(NB_TRIPS = unlist(compiled_TRIPID_values),
PER_ID = unlist(compiled_PER_ID_values)) # shows number of trips by person IDs
Tbl_Multi_Transport <- merge(Tbl_Multi_Transport, SEQUENCE_TR, by = "PER_ID",
all.x = TRUE) # add all info in SEQUENCE to previous table (match by Person
IDs)
TYPE_NOT_CONSIDERED <- c("Airplane", "Walk") # list of type of transports not
considered in the multimodal test
TYPE_PERSO_CARS <- c("Car","Van", "SUV/Crossover", "Pickup truck",
"Recreational Vehicle") # list of personal vehicles or cars

```

```

TYPE_PUBLIC_TRANSPORT <- c("Public or commuter bus", "School bus", "Street car
or trolley car", "Subway or elevated rail", "Commuter rail", "Amtrak") # list
of public transports
TYPE_DIVERS <- c("Motorcycle", "Taxicab or limo service", "Other ride-sharing
service", "Paratransit/ Dial a ride", "Bicycle (including bikeshare,
ebike...)", "E-scooter", "Other (specify)") #list of other modes
occurrence_2 = 0 # initial value to 0 (used to count total number of
multimodals)
compiled_multimodal_ID <- list() # list of person multimodal by ID (empty)
compiled_multimodal_SEQ <- list() # list of trip sequence that are multimodal
(empty)

for (j in 1:nrow(Tbl_Multi_Transport)) { # for all rows in table with the
lists of all transport sequence by persons
  split_elements <- strsplit(Tbl_Multi_Transport$Vehicle_Sequence[j], ",
")[[1]] # split the sequence thanks to "," character
  if (!any(split_elements %in% TYPE_NOT_CONSIDERED)){ # if there is not any
element from "Not considered" category
    if (any(split_elements %in% TYPE_PUBLIC_TRANSPORT) || any(split_elements
%in% TYPE_DIVERS)){ # if there is at least one element from "public
transport" of "divers" in the sequence
      if (length(split_elements) > 1) { # if there are at least 2 elements
in the sequence of transports
        for (k in 1:(length(split_elements) - 1)) { # for each element in
the sequence of transports
          if (!is.na(split_elements[k]) && !is.na(split_elements[k + 1]) &&
split_elements[k] != split_elements[k+1]) { # if two consecutive elements are
different
            compiled_multimodal_ID <- append(compiled_multimodal_ID,
Tbl_Multi_Transport$PER_ID[j]) # put the ID of the person that matches all
conditions in the list
            compiled_multimodal_SEQ <- append(compiled_multimodal_SEQ,
Tbl_Multi_Transport$Vehicle_Sequence[j]) # put the sequence of transports of
the person that matches all conditions in the list
            occurrence_2 <- occurrence_2 + 1 # count the number of elements
in the sequence
            break # stop the loop when it is done
          }
        }
      }
    }
  }
}

library(knitr) # Charging library
library(kableExtra) # Charging library
Tbl_Multimodal <- data.frame() # creating a table with all the persons
multimodal (empty)
Tbl_Multimodal <- data.frame(PER_ID = unlist(compiled_multimodal_ID) , # add
the IDs of all the persons multimodal in the first column

```

```

        compiled_multimodal_SEQ = unlist(compiled_multimodal_SEQ)) #
add sequence of transport for each person multimodal
Percent_Multimodal <- 100*(occurrence_2/occurrence_1) # calculates the
percentage of people that are multimodal
Tbl_Multimodal <- merge(Tbl_Multimodal, Tbl_Multi_Transport, by = "PER_ID",
all.x = TRUE) # add all the info from the general table of transports to the
new table (match by person IDs)
Tbl_Multimodal <- Tbl_Multimodal[,-2] # delete 2 last column that are
redundant
colnames(Tbl_Multimodal)[which(colnames(Tbl_Multimodal) ==
"Vehicle_Sequence")] <- "Vehicle_Sequence_Multimodal" # changing name of
column "Vehicle_Sequence_Multimodal"
max_Value <- max(Tbl_Multimodal$NB_TRIPS) # calculates the maximum number of
trips in the multimodal table
max_values_table <- data.frame("Valeur maximale" = max_Value, "PER_ID" =
Tbl_Multi_Transport$PER_ID[Tbl_Multi_Transport$NB_TRIPS == max_Value]) # recap
of all previous calculations and ratios
Superior_at <- 1 # condition to only see the people with at least 2 different
transportation modes
TRIPID_Sup_Value <- Tbl_Multimodal[Tbl_Multimodal$NB_TRIPS > Superior_at,] #
if number of trips is superior to the reference value then compile
Lengh_TRIPID_Sup_Value <- length(TRIPID_Sup_Value$NB_TRIPS) # calculates the
number of rows in TRIP_Sup_Value

library(dplyr) # charging library
data_TRIP_Person <- data.frame() # table that concatenates previous table
with many data_TRIP's columns but for each person
data_TRIP_Person <- merge(Tbl_Multi_Transport[, c("PER_ID", "NB_TRIPS",
"Trip_Sequence", "Vehicle_Sequence")],
                        Tbl_Multimodal[,c("PER_ID",
"Vehicle_Sequence_Multimodal")], by = "PER_ID", all.x = TRUE) # adds columns
to original data source to the table created by
data_TRIP_Person <- merge(data_TRIP_Person[,],
Tbl_Multi_Transport[,c("PER_ID", "VEHCASEID", "PUBTRANS", "TRPMILES",
"STRTTIME", "ENDTIME", "DWELTIME", "TRVLCMIN")], by
= "PER_ID", all.x = TRUE) # adds columns to original data source to the
table created by person IDs
data_TRIP_Person$Multi <- NA # makes sure there are no values left in "Multi"
column
subset_data_TRIP_columns <- data_TRIP[, c("PER_ID",
"MULTIMODAL", "TDAYDATE", "DATE", "TRAVDAY", "DAY", "DRIVER", "DRIVER_STATUS", "LIF_C
YC", "FAMILY", "R_AGE", "R_SEX", "SEXE", "EDUC",
"EDUCATION", "HHFAMINC", "HH_INCOME", "
R_RACE", "RACE", "R_HISP", "HISP", "WRKCOUNT", "WORKER", "EMPLOYMENT", "URBAN", "URBN"
, "URBANSIZE",
"DENSITY", "URBRUR", "URBANIZED", "WHYT
O", "PURPOSE", "TDWKND", "TIME_WEEK", "PRMACT", "ACTIVITY", "DRVRCNT", "HHVEHCNT", "NU
MADLT")] # takes key variables from original data source

```

```

unique_subset_data <-
subset_data_TRIP_columns[!duplicated(subset_data_TRIP_columns$PER_ID), ] #
intermediary table to do the next merge by keeping persons IDs
data_TRIP_Person <- merge(data_TRIP_Person, unique_subset_data, by = "PER_ID",
all.x = TRUE) # add all previous categories selected form data source and the
previous tables in a new one
data_TRIP_Person_nrows <- nrow(data_TRIP_Person) # number of row of table of
persons (first version)
for (l in 1:data_TRIP_Person_nrows) { #max : total number of persons
  if (any(!is.na(data_TRIP_Person$Vehicle_Sequence_Multimodal[l]) &
data_TRIP_Person$Vehicle_Sequence_Multimodal[l] != "")) { # is there is a
multimodal sequence
    data_TRIP_Person$MULTIMODAL[l] <- "YES" # add "YES" in the "Multimodal"
column
  } else { # otherwise...
    data_TRIP_Person$MULTIMODAL[l] <- "NO" # add "NO" in the "Multimodal"
column
  }
}
data_TRIP_Person$Multi[data_TRIP_Person$MULTIMODAL == "YES"] <- 1 # "Multi"
is 1 one for YES (factor)
data_TRIP_Person$Multi[data_TRIP_Person$MULTIMODAL == "NO"] <- 0 # "Multi"
is one for NO (factor)

## >>> AGE GROUPS :
library(reshape2) # charging library
age_breaks <- c(5, 13, 16, 25, 35, 50, 65, 93) # boundaries for the age
groups based on data (from 5 to 92 + 1 years old for the last one)
age.group <- cut(as.numeric(data_TRIP_Person$R_AGE), breaks = age_breaks) #
creation of the intervals based on the boundaries
library(dplyr) # charging library (python function)
data_TRIP_Person <- data_TRIP_Person %>%
  mutate(Age_Intervals = NA, .after = R_AGE) # put the new column with
intervals after the Age column
data_TRIP_Person$Age_Intervals <- NA #creation of new column for the intervals
for (p in 1:nrow(data_TRIP_Person)) { # read data_TRIP_Person row by row
  if (data_TRIP_Person$R_AGE[p] >= age_breaks[1] && data_TRIP_Person$R_AGE[p]
< age_breaks[2]) { # if the age is between the first and second boundaries...
    data_TRIP_Person$Age_Intervals[p] <- "(5,13[" # then indicates that
intervals in the column with age intervals
  } else if (data_TRIP_Person$R_AGE[p] >= age_breaks[2] &&
data_TRIP_Person$R_AGE[p] < age_breaks[3]) { # same for the other intervals...
    data_TRIP_Person$Age_Intervals[p] <- "(13,16["
  } else if (data_TRIP_Person$R_AGE[p] >= age_breaks[3] &&
data_TRIP_Person$R_AGE[p] < age_breaks[4]) {
    data_TRIP_Person$Age_Intervals[p] <- "(16,25["
  } else if (data_TRIP_Person$R_AGE[p] >= age_breaks[4] &&
data_TRIP_Person$R_AGE[p] < age_breaks[5]) {
    data_TRIP_Person$Age_Intervals[p] <- "(25,35["
  }
}

```

```

    } else if (data_TRIP_Person$R_AGE[p] >= age_breaks[5] &&
data_TRIP_Person$R_AGE[p] < age_breaks[6]) {
      data_TRIP_Person$Age_Intervals[p] <- "(35,50["
    } else if (data_TRIP_Person$R_AGE[p] >= age_breaks[6] &&
data_TRIP_Person$R_AGE[p] < age_breaks[7]) {
      data_TRIP_Person$Age_Intervals[p] <- "(50,65["
    } else if (data_TRIP_Person$R_AGE[p] >= age_breaks[7] &&
data_TRIP_Person$R_AGE[p] < age_breaks[8]) {
      data_TRIP_Person$Age_Intervals[p] <- "(65,92]"
    }
  }
}

## >>> DISTANCE GROUPS (same logic as age groups):
library(reshape2) # charging libraries
library(dplyr)
distance_breaks <- c(0,0.5, 2, 5, 10, 25, 50, 100, 5000) # distance
boundaries (in miles)
distance.group <- cut(as.numeric(data_TRIP_Person$TRPMILES), breaks =
distance_breaks) # distance intervals created based on boundaries
data_TRIP_Person <- data_TRIP_Person %>%
  mutate(Distance_Intervals = NA, .after = TRPMILES) # put new column with
distance intervals after original distance column
data_TRIP_Person$TRPMILES[is.na(data_TRIP_Person$TRPMILES)] <- 0 # makes sure
are values in the distance column, otherwise takes "0"
for (q in 1:nrow(data_TRIP_Person)) { # read data_TRIP_Person row by row
  if (data_TRIP_Person$TRPMILES[q] >= distance_breaks[1] &&
data_TRIP_Person$TRPMILES[q] < distance_breaks[2]) { # if the distance is
between first and second boundaries...
    data_TRIP_Person$Distance_Intervals[q] <- "(0,0.5[" # then indicates that
intervals in the column with distance intervals
  } else if (data_TRIP_Person$TRPMILES[q] >= distance_breaks[2] &&
data_TRIP_Person$TRPMILES[q] < distance_breaks[3]) { # same for the other
intervals...
    data_TRIP_Person$Distance_Intervals[q] <- "(0.5,2["
  } else if (data_TRIP_Person$TRPMILES[q] >= distance_breaks[3] &&
data_TRIP_Person$TRPMILES[q] < distance_breaks[4]) {
    data_TRIP_Person$Distance_Intervals[q] <- "(2,5["
  } else if (data_TRIP_Person$TRPMILES[q] >= distance_breaks[4] &&
data_TRIP_Person$TRPMILES[q] < distance_breaks[5]) {
    data_TRIP_Person$Distance_Intervals[q] <- "(5,10["
  } else if (data_TRIP_Person$TRPMILES[q] >= distance_breaks[5] &&
data_TRIP_Person$TRPMILES[q] < distance_breaks[6]) {
    data_TRIP_Person$Distance_Intervals[q] <- "(10,25["
  } else if (data_TRIP_Person$TRPMILES[q] >= distance_breaks[6] &&
data_TRIP_Person$TRPMILES[q] < distance_breaks[7]) {
    data_TRIP_Person$Distance_Intervals[q] <- "(25,50["
  } else if (data_TRIP_Person$TRPMILES[q] >= distance_breaks[7] &&
data_TRIP_Person$TRPMILES[q] < distance_breaks[8]) {
    data_TRIP_Person$Distance_Intervals[q] <- "(50,100["
  }
}

```

```

    } else if (data_TRIP_Person$TRPMILES[q] >= distance_breaks[8] &&
data_TRIP_Person$TRPMILES[q] < distance_breaks[9]) {
      data_TRIP_Person$Distance_Intervals[q] <- "(100,5000)"
    }
  }
}

## >>> TRAVEL TIME GROUPS (same logic as age or distance groups):
library(reshape2) # charging libraries
library(dplyr)
travel_time_breaks <- c(0,6, 11, 16, 21, 31, 46, 61,121,241,361,1441) # time
boundaries (in minutes)
travel_time.group <- cut(as.numeric(data_TRIP_Person$TRVLCMIN), breaks =
travel_time_breaks) # time intervals created based on boundaries
data_TRIP_Person <- data_TRIP_Person %>%
  mutate(Travel_time_Intervals = NA, .after = TRVLCMIN) # put new column with
time intervals after original travel time column
data_TRIP_Person$TRVLCMIN[is.na(data_TRIP_Person$TRVLCMIN)] <- 0 # makes sure
are values in the travel time column, otherwise takes "0"
for (r in 1:nrow(data_TRIP_Person)) { # read data_TRIP_Person row by row
  if (data_TRIP_Person$TRVLCMIN[r] >= travel_time_breaks[1] &&
data_TRIP_Person$TRVLCMIN[r] < travel_time_breaks[2]) { # if the travel time
is between first and second boundaries...
    data_TRIP_Person$Travel_time_Intervals[r] <- "(0,5[" # then indicates
that intervals in the column with time intervals
  } else if (data_TRIP_Person$TRVLCMIN[r] >= travel_time_breaks[2] &&
data_TRIP_Person$TRVLCMIN[r] < travel_time_breaks[3]) { # same for the other
intervals...
    data_TRIP_Person$Travel_time_Intervals[r] <- "(5,10["
  } else if (data_TRIP_Person$TRVLCMIN[r] >= travel_time_breaks[3] &&
data_TRIP_Person$TRVLCMIN[r] < travel_time_breaks[4]) {
    data_TRIP_Person$Travel_time_Intervals[r] <- "(10,15["
  } else if (data_TRIP_Person$TRVLCMIN[r] >= travel_time_breaks[4] &&
data_TRIP_Person$TRVLCMIN[r] < travel_time_breaks[5]) {
    data_TRIP_Person$Travel_time_Intervals[r] <- "(15,20["
  } else if (data_TRIP_Person$TRVLCMIN[r] >= travel_time_breaks[5] &&
data_TRIP_Person$TRVLCMIN[r] < travel_time_breaks[6]) {
    data_TRIP_Person$Travel_time_Intervals[r] <- "(20,30["
  } else if (data_TRIP_Person$TRVLCMIN[r] >= travel_time_breaks[6] &&
data_TRIP_Person$TRVLCMIN[r] < travel_time_breaks[7]) {
    data_TRIP_Person$Travel_time_Intervals[r] <- "(30,45["
  } else if (data_TRIP_Person$TRVLCMIN[r] >= travel_time_breaks[7] &&
data_TRIP_Person$TRVLCMIN[r] < travel_time_breaks[8]) {
    data_TRIP_Person$Travel_time_Intervals[r] <- "(45,60["
  } else if (data_TRIP_Person$TRVLCMIN[r] >= travel_time_breaks[8] &&
data_TRIP_Person$TRVLCMIN[r] < travel_time_breaks[9]) {
    data_TRIP_Person$Travel_time_Intervals[r] <- "(60,120["
  } else if (data_TRIP_Person$TRVLCMIN[r] >= travel_time_breaks[9] &&
data_TRIP_Person$TRVLCMIN[r] < travel_time_breaks[10]) {
    data_TRIP_Person$Travel_time_Intervals[r] <- "(120,240["
  }
}

```

```

    } else if (data_TRIP_Person$TRVLCMIN[r] >= travel_time_breaks[10] &&
data_TRIP_Person$TRVLCMIN[r] < travel_time_breaks[11]) {
    data_TRIP_Person$Travel_time_Intervals[r] <- "(240,360["
    } else if (data_TRIP_Person$TRVLCMIN[r] >= travel_time_breaks[11] &&
data_TRIP_Person$TRVLCMIN[r] < travel_time_breaks[12]) {
    data_TRIP_Person$Travel_time_Intervals[r] <- "(360,1140]"
    } else {
    data_TRIP_Person$Travel_time_Intervals[r] <- "(360,1140]"
    }
}
}

```

```

## Logic of the following lines : to concatenate existing categories in new
ones to have least categories in total per variable
## we make sure to have in each case 2 columns for each variable : one for the
name with details and one for the logistic after that (with numbers)

```

```

## >>> rework of the dataset : "LIF_CYC/FAMILY" (from 10 to 7 categories) :
data_TRIP_Person$FAMILY[data_TRIP_Person$FAMILY == "one adult, retired, no
children"] <- "one adult, no children" # example : retired and unique adults
are concatenated
data_TRIP_Person$FAMILY[data_TRIP_Person$FAMILY == "2+ adults, no children"]
<- "2+ adults, no children"
data_TRIP_Person$FAMILY[data_TRIP_Person$FAMILY == "2+ adults, retired, no
children"] <- "2+ adults, no children"
data_TRIP_Person$FAMILY[data_TRIP_Person$FAMILY == "one adult, youngest child
0-5"] <- "one adult, youngest child 0-15"
data_TRIP_Person$FAMILY[data_TRIP_Person$FAMILY == "one adult, youngest child
6-15"] <- "one adult, youngest child 0-15"
data_TRIP_Person$FAMILY[data_TRIP_Person$FAMILY == "2+ adults, youngest child
0-5"] <- "2+ adults, youngest child 0-15"
data_TRIP_Person$FAMILY[data_TRIP_Person$FAMILY == "2+ adults, youngest child
6-15"] <- "2+ adults, youngest child 0-15"
data_TRIP_Person$LIF_CYC[data_TRIP_Person$FAMILY == "one adult, no children"]
<- 1 # factor in the other column "LIF_CYC" corresponding to "FAMILY"
data_TRIP_Person$LIF_CYC[data_TRIP_Person$FAMILY == "2+ adults, no children"]
<- 2
data_TRIP_Person$LIF_CYC[data_TRIP_Person$FAMILY == "one adult, youngest child
0-15"] <- 3
data_TRIP_Person$LIF_CYC[data_TRIP_Person$FAMILY == "2+ adults, youngest child
0-15"] <- 4
data_TRIP_Person$LIF_CYC[data_TRIP_Person$FAMILY == "one adult, youngest child
16-21"] <- 5
data_TRIP_Person$LIF_CYC[data_TRIP_Person$FAMILY == "2+ adults, youngest child
16-21"] <- 6
data_TRIP_Person$LIF_CYC[data_TRIP_Person$FAMILY == "Other Family"] <- 7

```

```

## >>> rework of the dataset : "EDUC/EDUCATION" (from 8 to 7 categories) :
data_TRIP_Person$EDUCATION[data_TRIP_Person$EDUCATION == "Some high school, no
diploma/GED"] <- "High school"

```

```

data_TRIP_Person$EDUCATION[data_TRIP_Person$EDUCATION == "High school
graduate, have diploma/GED"] <- "High school"
data_TRIP_Person$EDUCATION[data_TRIP_Person$EDUCATION == "Some college, no
degree/ some trade school"] <- "College"
data_TRIP_Person$EDUCATION[data_TRIP_Person$EDUCATION == "Associates degree
(2-year)/ trade school certificate"] <- "College"
data_TRIP_Person$EDUCATION[is.na(data_TRIP_Person$EDUCATION)] <- "Other
Education"
data_TRIP_Person$EDUC[data_TRIP_Person$EDUCATION == "Less than high school"]
<- 1
data_TRIP_Person$EDUC[data_TRIP_Person$EDUCATION == "High school"] <- 2
data_TRIP_Person$EDUC[data_TRIP_Person$EDUCATION == "College"] <- 3
data_TRIP_Person$EDUC[data_TRIP_Person$EDUCATION == "Bachelor's degree"] <- 4
data_TRIP_Person$EDUC[data_TRIP_Person$EDUCATION == "Master's degree"] <- 5
data_TRIP_Person$EDUC[data_TRIP_Person$EDUCATION == "Professional/Doctorate
degree"] <- 6
data_TRIP_Person$EDUC[data_TRIP_Person$EDUCATION == "Other Education"] <- 7

## >>> rework of the dataset : "R_RACE/RACE" (from 7+1 to 5 categories) :
for (n in 1:nrow(data_TRIP_Person)) { # read data_TRIP_Person row by row
  if (data_TRIP_Person$R_HISP[n] == 1) { # if the person is hispanic (in a
    specific column "Hispanic")
      data_TRIP_Person$R_RACE[n] <- 3 # then we add this specific race variable
as a category in the "R_RACE" variable
      data_TRIP_Person$RACE[n] <- "Hispanic" #idem in the "RACE" column with the
names
    } else {}
  }
column_HISP_deleat <- c("R_HISP", "HISP") # columns to deletat by names
data_TRIP_Person <- data_TRIP_Person[, -which(names(data_TRIP_Person) %in%
column_HISP_deleat)] # columns linked to hispanic deleted
data_TRIP_Person$RACE[data_TRIP_Person$RACE == "American Indian/Alaska
Native"] <- "Other race"
data_TRIP_Person$RACE[data_TRIP_Person$RACE == "Multiple race selected"] <-
"Other race"
data_TRIP_Person$RACE[data_TRIP_Person$RACE == "Native Hawaiian/Pacific
Islander"] <- "Other race"
data_TRIP_Person$RACE[is.na(data_TRIP_Person$RACE)] <- "Other race" #
everything that is not already in a category goes to "Other race"
data_TRIP_Person$R_RACE[data_TRIP_Person$RACE == "Asian"] <- 1 # puts numbers
in order (factors) in "R_RACE" according to "RACE" columns
data_TRIP_Person$R_RACE[data_TRIP_Person$RACE == "Black or African
American "] <- 2
data_TRIP_Person$R_RACE[data_TRIP_Person$RACE == "Hispanic"] <- 3
data_TRIP_Person$R_RACE[data_TRIP_Person$RACE == "White"] <- 4
data_TRIP_Person$R_RACE[data_TRIP_Person$RACE == "Other race"] <- 5

## >>> rework of the dataset : "R_SEX/SEXE" (from 4 to 3 categories) :
data_TRIP_Person$SEXE[data_TRIP_Person$SEXE == "Don't know"] <- "Other Sex"

```

```

data_TRIP_Person$SEXE[data_TRIP_Person$SEXE == "Refuse"] <- "Other Sex"
data_TRIP_Person$R_SEX[data_TRIP_Person$SEXE == "Female"] <- 1
data_TRIP_Person$R_SEX[data_TRIP_Person$SEXE == "Male"] <- 2
data_TRIP_Person$R_SEX[data_TRIP_Person$SEXE == "Other Sex"] <- 3

## >>> rework of the dataset : "HHFAMINC/HH_INCOME" (from 13 to 4 categories)
:
data_TRIP_Person$HH_INCOME[data_TRIP_Person$HH_INCOME == "Less than $10,000"]
<- "Less than $24,999" # low revenues mixed
data_TRIP_Person$HH_INCOME[data_TRIP_Person$HH_INCOME == "$10,000 to $14,999"]
<- "Less than $24,999"
data_TRIP_Person$HH_INCOME[data_TRIP_Person$HH_INCOME == "$15,000 to $24,999"]
<- "Less than $24,999"
data_TRIP_Person$HH_INCOME[data_TRIP_Person$HH_INCOME == "$25,000 to $34,999"]
<- "$25,000 to $100,000" # medium revenues mixed
data_TRIP_Person$HH_INCOME[data_TRIP_Person$HH_INCOME == "$35,000 to $49,999"]
<- "$25,000 to $100,000"
data_TRIP_Person$HH_INCOME[data_TRIP_Person$HH_INCOME == "$50,000 to $74,999"]
<- "$25,000 to $100,000"
data_TRIP_Person$HH_INCOME[data_TRIP_Person$HH_INCOME == "$75,000 to $99,999"]
<- "$25,000 to $100,000"
data_TRIP_Person$HH_INCOME[data_TRIP_Person$HH_INCOME == "$100,000 to
$124,999"] <- "$100,000 or more" # large revenues mixed
data_TRIP_Person$HH_INCOME[data_TRIP_Person$HH_INCOME == "$125,000 to
$149,999"] <- "$100,000 or more"
data_TRIP_Person$HH_INCOME[data_TRIP_Person$HH_INCOME == "$150,000 to
$199,999"] <- "$100,000 or more"
data_TRIP_Person$HH_INCOME[data_TRIP_Person$HH_INCOME == "$200,000 or more"]
<- "$100,000 or more"
data_TRIP_Person$HH_INCOME[data_TRIP_Person$HH_INCOME == "I prefer not to
answer"] <- "Other Income"
data_TRIP_Person$HH_INCOME[data_TRIP_Person$HH_INCOME == "I don't know"] <-
"Other Income"
data_TRIP_Person$HHFAMINC[data_TRIP_Person$HH_INCOME == "Less than $24,999"]
<- 1
data_TRIP_Person$HHFAMINC[data_TRIP_Person$HH_INCOME == "$25,000 to $100,000"]
<- 2
data_TRIP_Person$HHFAMINC[data_TRIP_Person$HH_INCOME == "$100,000 or more"] <-
3
data_TRIP_Person$HHFAMINC[data_TRIP_Person$HH_INCOME == "Other Income"] <- 4

## >>> rework of the dataset : "URBANSIZE/DENSITY" (from 6 to 4 categories) :
data_TRIP_Person$DENSITY[data_TRIP_Person$DENSITY == "50,000-199,999"] <-
"50,000-499,000"
data_TRIP_Person$DENSITY[data_TRIP_Person$DENSITY == "200,000-499,999"] <-
"50,000-499,000"
data_TRIP_Person$DENSITY[data_TRIP_Person$DENSITY == "1,000,000 or more with
heavy rail"] <- "1,000,000 or more"

```

```

data_TRIP_Person$DENSITY[data_TRIP_Person$DENSITY == "1,000,000 or more
without heavy rail"] <- "1,000,000 or more"
data_TRIP_Person$DENSITY[data_TRIP_Person$DENSITY == "Not in urbanized area"]
<- "Other Density"
data_TRIP_Person$URBANSIZE[data_TRIP_Person$DENSITY == "50,000-499,000"] <- 1
data_TRIP_Person$URBANSIZE[data_TRIP_Person$DENSITY == "500,000-999,999"] <- 2
data_TRIP_Person$URBANSIZE[data_TRIP_Person$DENSITY == "1,000,000 or more"] <-
3
data_TRIP_Person$URBANSIZE[data_TRIP_Person$DENSITY == "Other Density"] <- 4

## >>> rework of the dataset : "WHYTO/PURPOSE" (from 20 to 10 categories) :
data_TRIP_Person$PURPOSE[data_TRIP_Person$PURPOSE == "Work from home (paid)"]
<- "Work"
data_TRIP_Person$PURPOSE[data_TRIP_Person$PURPOSE == "Work at a non-home
location"] <- "Work"
data_TRIP_Person$PURPOSE[data_TRIP_Person$PURPOSE == "Work activity to drop-
off/pickup someone/something"] <- "Work"
data_TRIP_Person$PURPOSE[data_TRIP_Person$PURPOSE == "Other work-related
activities"] <- "Work"
data_TRIP_Person$PURPOSE[data_TRIP_Person$PURPOSE == "Volunteer activities
(not paid)"] <- "Work"
data_TRIP_Person$PURPOSE[data_TRIP_Person$PURPOSE == "Attend childcare or
adult care"] <- "Health"
data_TRIP_Person$PURPOSE[data_TRIP_Person$PURPOSE == "Health care visit"] <-
"Health"
data_TRIP_Person$PURPOSE[data_TRIP_Person$PURPOSE == "Buy meals"] <-
"Spending"
data_TRIP_Person$PURPOSE[data_TRIP_Person$PURPOSE == "Shop/buy/pick-up or
return goods"] <- "Spending"
data_TRIP_Person$PURPOSE[data_TRIP_Person$PURPOSE == "Other family/personal
errands"] <- "Spending"
data_TRIP_Person$PURPOSE[data_TRIP_Person$PURPOSE == "Recreational
activities"] <- "Recreational"
data_TRIP_Person$PURPOSE[data_TRIP_Person$PURPOSE == "Religious or other
community activities"] <- "Recreational"
data_TRIP_Person$PURPOSE[data_TRIP_Person$PURPOSE == "Exercise"] <-
"Recreational"
data_TRIP_Person$PURPOSE[data_TRIP_Person$PURPOSE == "Rest or
relaxation/vacation"] <- "Recreational"
data_TRIP_Person$PURPOSE[data_TRIP_Person$PURPOSE == "Something else
(specify)"] <- "Other Purpose"
data_TRIP_Person$WHYTO[data_TRIP_Person$PURPOSE == "Regular activities at
home"] <- 1
data_TRIP_Person$WHYTO[data_TRIP_Person$PURPOSE == "Work"] <- 2
data_TRIP_Person$WHYTO[data_TRIP_Person$PURPOSE == "Attend school as a
student"] <- 3
data_TRIP_Person$WHYTO[data_TRIP_Person$PURPOSE == "Health"] <- 4
data_TRIP_Person$WHYTO[data_TRIP_Person$PURPOSE == "Change type of
transportation"] <- 5

```

```

data_TRIP_Person$WHYTO[data_TRIP_Person$PURPOSE == "Drop off/pick up someone
(personal)"] <- 6
data_TRIP_Person$WHYTO[data_TRIP_Person$PURPOSE == "Spending"] <- 7
data_TRIP_Person$WHYTO[data_TRIP_Person$PURPOSE == "Recreational"] <- 8
data_TRIP_Person$WHYTO[data_TRIP_Person$PURPOSE == "Visit friends or
relatives"] <- 9
data_TRIP_Person$WHYTO[data_TRIP_Person$PURPOSE == "Other Purpose"] <- 10

## >>> rework of the dataset : "PRMACT/ACTIVITY" (from 6+1 to 6 categories) :
data_TRIP_Person$ACTIVITY[data_TRIP_Person$ACTIVITY == "Temporarily absent
from a job or business"] <- "Unemployed"
data_TRIP_Person$ACTIVITY[data_TRIP_Person$ACTIVITY == "Looking for
work/unemployed"] <- "Unemployed"
data_TRIP_Person$ACTIVITY[data_TRIP_Person$ACTIVITY == "Something else"] <-
"Other Activity"
for (o in 1:nrow(data_TRIP_Person)) { # read data_TRIP_Person row by row
  if (is.na(data_TRIP_Person$WORKER[o])) { # if there are no values in
"WORKER" column...
    data_TRIP_Person$WORKER[o] <- 3 # indicates 3
  } else if (data_TRIP_Person$WORKER[o] == 1) { # if there is a "1"...
    data_TRIP_Person$PRMACT[o] <- 6
    data_TRIP_Person$ACTIVITY[o] <- "Worker" # transform worker variable into
a category in "ACTIVITY" (same case as "Hispanics" previously)
  } else {}
}
column_Worker_deleat <- c("WORKER", "EMPLOYMENT") # columns to delet
data_TRIP_Person <- data_TRIP_Person[, -which(names(data_TRIP_Person) %in%
column_Worker_deleat)] # columns linked to "WORKER" are deleted
data_TRIP_Person$ACTIVITY[is.na(data_TRIP_Person$ACTIVITY)] <- "Other
Activity"
data_TRIP_Person$PRMACT[data_TRIP_Person$ACTIVITY == "Unemployed"] <- 1
data_TRIP_Person$PRMACT[data_TRIP_Person$ACTIVITY == "A homemaker"] <- 2
data_TRIP_Person$PRMACT[data_TRIP_Person$ACTIVITY == "Going to school"] <- 3
data_TRIP_Person$PRMACT[data_TRIP_Person$ACTIVITY == "Retired"] <- 4
data_TRIP_Person$PRMACT[data_TRIP_Person$ACTIVITY == "Worker"] <- 5

## >>> rework of the dataset : "DRIVER/DRIVER_STATUS"
data_TRIP_Person$DRIVER[data_TRIP_Person$DRIVER_STATUS == "Not Driver"] <- 2 #
original "0" value is changed into "1" to be then factored (regression)

## >>> rework of the dataset : "DRIVERS/DRVRCNT"
library(reshape2)
library(dplyr)
data_TRIP_Person <- data_TRIP_Person %>%
  mutate(DRIVERS = NA, .after = ACTIVITY)
data_TRIP_Person$DRVRCNT[data_TRIP_Person$DRVRCNT == 4] <- "4+"
data_TRIP_Person$DRVRCNT[data_TRIP_Person$DRVRCNT == 5] <- "4+"
data_TRIP_Person$DRVRCNT[data_TRIP_Person$DRVRCNT == 6] <- "4+"
data_TRIP_Person$DRVRCNT[data_TRIP_Person$DRVRCNT == 7] <- "4+"

```

```

data_TRIP_Person$DRIVERS[data_TRIP_Person$DRVCNT == 0] <- 0
data_TRIP_Person$DRIVERS[data_TRIP_Person$DRVCNT == 1] <- 1
data_TRIP_Person$DRIVERS[data_TRIP_Person$DRVCNT == 2] <- 2
data_TRIP_Person$DRIVERS[data_TRIP_Person$DRVCNT == 3] <- 3
data_TRIP_Person$DRIVERS[data_TRIP_Person$DRVCNT == "4+"] <- 4

## >>> rework of the dataset : "HOUSEVEH/HHVEHCNT"
library(reshape2)
library(dplyr)
data_TRIP_Person <- data_TRIP_Person %>%
  mutate(HOUSEVEH = NA, .after = DRVCNT)
data_TRIP_Person$HHVEHCNT[data_TRIP_Person$HHVEHCNT == 4] <- "4+"
data_TRIP_Person$HHVEHCNT[data_TRIP_Person$HHVEHCNT == 5] <- "4+"
data_TRIP_Person$HHVEHCNT[data_TRIP_Person$HHVEHCNT == 6] <- "4+"
data_TRIP_Person$HHVEHCNT[data_TRIP_Person$HHVEHCNT == 7] <- "4+"
data_TRIP_Person$HHVEHCNT[data_TRIP_Person$HHVEHCNT == 8] <- "4+"
data_TRIP_Person$HHVEHCNT[data_TRIP_Person$HHVEHCNT == 9] <- "4+"
data_TRIP_Person$HHVEHCNT[data_TRIP_Person$HHVEHCNT == 10] <- "4+"
data_TRIP_Person$HHVEHCNT[data_TRIP_Person$HHVEHCNT == 11] <- "4+"
data_TRIP_Person$HOUSEVEH[data_TRIP_Person$HHVEHCNT == 0] <- 0
data_TRIP_Person$HOUSEVEH[data_TRIP_Person$HHVEHCNT == 1] <- 1
data_TRIP_Person$HOUSEVEH[data_TRIP_Person$HHVEHCNT == 2] <- 2
data_TRIP_Person$HOUSEVEH[data_TRIP_Person$HHVEHCNT == 3] <- 3
data_TRIP_Person$HOUSEVEH[data_TRIP_Person$HHVEHCNT == "4+"] <- 4

## >>> rework of the dataset : "ADULTS/NUMADLT"
library(reshape2)
library(dplyr)
data_TRIP_Person <- data_TRIP_Person %>%
  mutate(ADULTS = NA, .after = HHVEHCNT)
data_TRIP_Person$NUMADLT[data_TRIP_Person$NUMADLT == 4] <- "4+"
data_TRIP_Person$NUMADLT[data_TRIP_Person$NUMADLT == 5] <- "4+"
data_TRIP_Person$NUMADLT[data_TRIP_Person$NUMADLT == 6] <- "4+"
data_TRIP_Person$NUMADLT[data_TRIP_Person$NUMADLT == 7] <- "4+"
data_TRIP_Person$NUMADLT[data_TRIP_Person$NUMADLT == 8] <- "4+"
data_TRIP_Person$ADULTS[data_TRIP_Person$NUMADLT == 0] <- 0
data_TRIP_Person$ADULTS[data_TRIP_Person$NUMADLT == 1] <- 1
data_TRIP_Person$ADULTS[data_TRIP_Person$NUMADLT == 2] <- 2
data_TRIP_Person$ADULTS[data_TRIP_Person$NUMADLT == 3] <- 3
data_TRIP_Person$ADULTS[data_TRIP_Person$NUMADLT == "4+"] <- 4

## >>> COPY of dataset :
data_TRIP_Person_V2 <- data_TRIP_Person # copy of the person table

## >>> MORE CONTRACTIONS (specific to logistic regression) :
## rework of the dataset : "WHYTO/PURPOSE" for the regression (from 10
categories to 5) :
data_TRIP_Person_V2$PURPOSE[data_TRIP_Person_V2$PURPOSE == "Work"] <- "Other
Mandatory"

```

```

data_TRIP_Person_V2$PURPOSE[data_TRIP_Person_V2$PURPOSE == "Health"] <- "Other
Mandatory"
data_TRIP_Person_V2$PURPOSE[data_TRIP_Person_V2$PURPOSE == "Drop off/pick up
someone (personal)"] <- "Other Mandatory"
data_TRIP_Person_V2$PURPOSE[data_TRIP_Person_V2$PURPOSE == "Spending"] <-
"Non mandatory"
data_TRIP_Person_V2$PURPOSE[data_TRIP_Person_V2$PURPOSE == "Recreational"] <-
"Non mandatory"
data_TRIP_Person_V2$PURPOSE[data_TRIP_Person_V2$PURPOSE == "Visit friends or
relatives"] <- "Non mandatory"
data_TRIP_Person_V2$PURPOSE[data_TRIP_Person_V2$PURPOSE == "Other Purpose"] <-
"Non mandatory"
data_TRIP_Person_V2$WHYTO <- NA
data_TRIP_Person_V2$WHYTO[data_TRIP_Person_V2$PURPOSE == "Regular activities
at home"] <- 1
data_TRIP_Person_V2$WHYTO[data_TRIP_Person_V2$PURPOSE == "Other Mandatory"] <-
2
data_TRIP_Person_V2$WHYTO[data_TRIP_Person_V2$PURPOSE == "Attend school as a
student"] <- 3
data_TRIP_Person_V2$WHYTO[data_TRIP_Person_V2$PURPOSE == "Non mandatory"] <- 4
data_TRIP_Person_V2$WHYTO[data_TRIP_Person_V2$PURPOSE == "Change type of
transportation"] <- 5

## rework of the dataset : "Age_Intervals" for the regression (from 5
categories to 4) :
data_TRIP_Person_V2$Age_Intervals[data_TRIP_Person_V2$Age_Intervals ==
"(5,13["] <- "(5,16]"
data_TRIP_Person_V2$Age_Intervals[data_TRIP_Person_V2$Age_Intervals ==
"(13,16["] <- "(5,16]"
data_TRIP_Person_V2$Age_Intervals[data_TRIP_Person_V2$Age_Intervals ==
"(16,25["] <- "(16,35]"
data_TRIP_Person_V2$Age_Intervals[data_TRIP_Person_V2$Age_Intervals ==
"(25,35["] <- "(16,35]"
data_TRIP_Person_V2$Age_Intervals[data_TRIP_Person_V2$Age_Intervals ==
"(50,65["] <- "(50,92]"
data_TRIP_Person_V2$Age_Intervals[data_TRIP_Person_V2$Age_Intervals ==
"(65,92["] <- "(50,92]"

## rework of the dataset : "Travel_time_Intervals" for the regression (from 6
categories to 4) :
data_TRIP_Person_V2$Travel_time_Intervals[data_TRIP_Person_V2$Travel_time_Interval
s == "(0,5["] <- "(0,30["
data_TRIP_Person_V2$Travel_time_Intervals[data_TRIP_Person_V2$Travel_time_Interval
s == "(5,10["] <- "(0,30["
data_TRIP_Person_V2$Travel_time_Intervals[data_TRIP_Person_V2$Travel_time_Interval
s == "(10,15["] <- "(0,30["
data_TRIP_Person_V2$Travel_time_Intervals[data_TRIP_Person_V2$Travel_time_Interval
s == "(15,20["] <- "(0,30["

```

```

data_TRIP_Person_V2$Travel_time_Intervals[data_TRIP_Person_V2$Travel_time_Inte
rvals == "(20,30["] <- "(0,30["
data_TRIP_Person_V2$Travel_time_Intervals[data_TRIP_Person_V2$Travel_time_Inte
rvals == "(30,45["] <- "(30,60["
data_TRIP_Person_V2$Travel_time_Intervals[data_TRIP_Person_V2$Travel_time_Inte
rvals == "(45,60["] <- "(30,60["
data_TRIP_Person_V2$Travel_time_Intervals[data_TRIP_Person_V2$Travel_time_Inte
rvals == "(120,240["] <- "(120,1140["
data_TRIP_Person_V2$Travel_time_Intervals[data_TRIP_Person_V2$Travel_time_Inte
rvals == "(240,360["] <- "(120,1140["
data_TRIP_Person_V2$Travel_time_Intervals[data_TRIP_Person_V2$Travel_time_Inte
rvals == "(360,1140["] <- "(120,1140["
data_TRIP_Person_V2$Travel_time_Intervals[data_TRIP_Person_V2$Travel_time_Inte
rvals == "(360,1140["] <- "(120,1140["

```

```

## rework of the dataset : "Distance_Intervals" for the regression (from 5
categories to 4) :

```

```

data_TRIP_Person_V2$Distance_Intervals[data_TRIP_Person_V2$Distance_Intervals
== "(0,0.5["] <- "(0,10["
data_TRIP_Person_V2$Distance_Intervals[data_TRIP_Person_V2$Distance_Intervals
== "(0.5,2["] <- "(0,10["
data_TRIP_Person_V2$Distance_Intervals[data_TRIP_Person_V2$Distance_Intervals
== "(2,5["] <- "(0,10["
data_TRIP_Person_V2$Distance_Intervals[data_TRIP_Person_V2$Distance_Intervals
== "(5,10["] <- "(0,10["
data_TRIP_Person_V2$Distance_Intervals[data_TRIP_Person_V2$Distance_Intervals
== "(50,100["] <- "(50,5000["
data_TRIP_Person_V2$Distance_Intervals[data_TRIP_Person_V2$Distance_Intervals
== "(100,5000["] <- "(50,5000["
data_TRIP_Person_V2$Distance_Intervals[data_TRIP_Person_V2$Distance_Intervals
== "(100,5000["] <- "(50,5000["

```

```

## >>> FINAL table with data from original data trip + data trip person :

```

```

data_TRIP_V2 <- data_TRIP # Copy of original dataset
library(dplyr) # charging libraries
data_TRIP_V2 <- data_TRIP_V2 %>% # python based function
  mutate_all(~ ifelse(is.na(.), "NA", .)) # replace all empty values ("NA")
by the text value "NA" before the merge
data_TRIP_Final <- data.frame(data_TRIP_V2[, c("TRIPID", "TDCASEID",
"PERSONID", "PER_ID", "HOUSEID", "LOOP_TRIP")])
data_TRIP_Final <- merge(data_TRIP_Final, data_TRIP_Person_V2[, c("PER_ID",
"NB_TRIPS", "Trip_Sequence")], by = "PER_ID", all.x = TRUE)
library(dplyr) # charging libraries
data_TRIP_Final <- data_TRIP_Final %>% # python based function
  relocate(NB_TRIPS, .after = TRIPID) %>% # following lines : to put some
columns after another to reorganize the whole table
  relocate(Trip_Sequence, .after = NB_TRIPS)

```

```

data_TRIP_Final <-merge(data_TRIP_Final, data_TRIP_V2[,c("TDCASEID",
"TRPTRANS", "TR_TYPE")], by = "TDCASEID", all.x = TRUE)
data_TRIP_Final <-merge(data_TRIP_Final, data_TRIP_Person_V2[,c("PER_ID",
"Vehicle_Sequence")], by = "PER_ID", all.x = TRUE)
data_TRIP_Final <-merge(data_TRIP_Final, data_TRIP_V2[,c("TDCASEID",
"VEHCASEID", "VEH_TYPE", "VEH_TYPE")], by = "TDCASEID", all.x = TRUE)
data_TRIP_Final <-merge(data_TRIP_Final, data_TRIP_V2[,c("TDCASEID",
"PUBTRANS")], by = "TDCASEID", all.x = TRUE)

data_TRIP_Final <-merge(data_TRIP_Final,
data_TRIP_Person_V2[,c("PER_ID", "STRTTIME", "ENDTIME", "DWELTIME",
"TRVLCMIN", "Travel_time_Intervals", "TRPMILES", "Distance_Intervals")], by =
"PER_ID", all.x = TRUE)
names(data_TRIP_Final)[which(names(data_TRIP_Final) == "STRTTIME")] <-
"TOT_STRTTIME" # renaming the following columns to differentiate the ones for
the persons from the ones for the trip
names(data_TRIP_Final)[which(names(data_TRIP_Final) == "ENDTIME")] <-
"TOT_ENDTIME"
names(data_TRIP_Final)[which(names(data_TRIP_Final) == "DWELTIME")] <-
"TOT_DWELTIME"
names(data_TRIP_Final)[which(names(data_TRIP_Final) == "TRVLCMIN")] <-
"TOT_TRVLCMIN"
names(data_TRIP_Final)[which(names(data_TRIP_Final) ==
"Travel_time_Intervals")] <- "TOT_Travel_time_Intervals"
names(data_TRIP_Final)[which(names(data_TRIP_Final) == "TRPMILES")] <-
"TOT_TRPMILES"
names(data_TRIP_Final)[which(names(data_TRIP_Final) == "Distance_Intervals")]
<- "TOT_Distance_Intervals"

data_TRIP_Final <-merge(data_TRIP_Final, data_TRIP_V2[,c("TDCASEID",
"TDAYDATE", "DATE", "TRAVDAY", "DAY",
"STRTTIME", "ENDTIME", "DWELTIME", "TRVLCMIN", "TRPMILES")], by = "TDCASEID",
all.x = TRUE)
library(dplyr) # charging libraries
data_TRIP_Final <- data_TRIP_Final %>% # python based function
  relocate(TDAYDATE, .after = PUBTRANS) %>% # following lines : to put some
columns after another to reorganize the whole table
  relocate(DATE, .after = TDAYDATE) %>%
  relocate(TRAVDAY, .after = DATE) %>%
  relocate(DAY, .after = TRAVDAY) %>%
  relocate(STRTTIME, .after = DAY) %>%
  relocate(ENDTIME, .after = STRTTIME) %>%
  relocate(DWELTIME, .after = TOT_ENDTIME) %>%
  relocate(TRVLCMIN, .after = TOT_DWELTIME) %>%
  relocate(TRPMILES, .after = TOT_Travel_time_Intervals)
data_TRIP_Final <-merge(data_TRIP_Final, data_TRIP_Person_V2[,c("PER_ID",
"Multi", "MULTIMODAL", "DRIVER", "DRIVER_STATUS", "LIF_CYC", "FAMILY", "R_AGE", "Age_
Intervals", "R_SEX", "SEXE", "EDUC",

```

```

"EDUCATION", "H
HFAMINC", "HH_INCOME", "R_RACE", "RACE", "WRKCOUNT", "URBAN", "URBN", "URBANSIZE", "DE
NSITY", "URBRUR", "URBANIZED", "WHYTO", "PURPOSE", "TDWKND", "TIME_WEEK",
"PRMACT", "ACTI
VITY", "DRVRCNT", "HHVEHCNT", "NUMADLT")], by = "PER_ID", all.x = TRUE)
names(data_TRIP_Final)[which(names(data_TRIP_Final) == "Multi")] <-
"Multi_Persons"
names(data_TRIP_Final)[which(names(data_TRIP_Final) == "MULTIMODAL")] <-
"MULTIMODAL_Persons"
library(dplyr) # charging libraries
data_TRIP_Final <- data_TRIP_Final %>% # python based function
  relocate(Multi_Persons, .after = HOUSEID) %>% # following lines : to put
some columns after another to reorganize the whole table
  relocate(MULTIMODAL_Persons, .after = Multi_Persons)
data_TRIP_Final <- merge(data_TRIP_Final, data_TRIP_V2[,c("TDCASEID",
"TRPHHVEH", "NUMONTRP", "HHMEMDRV", "NONHHCNT", "HHACCCNT", "WHODROVE",
"DRVR_FLG", "PSGR_FLG", "ONTD_P1", "ONTD_P2", "ONTD_P3", "ONTD_P4", "ONTD_P5",
"ONTD_P6", "ONTD_P7", "ONTD_P8", "ONTD_P9", "ONTD_P10")], by = "TDCASEID",
all.x = TRUE)

## >>> Distance categories for each trip ("déplacement") :
library(dplyr) # charging libraries
distance_breaks_trip <- c(0,0.5, 2, 5, 10, 25, 50, 100, 5000) # defining the
limits for the intervals in miles
distance.group_trip <- cut(as.numeric(data_TRIP_Final$TRPMILES), breaks =
distance_breaks_trip) # defining the intervals
data_TRIP_Final <- data_TRIP_Final %>% # python based function
  mutate(Distance_Intervals_Trip = NA, .after = TRPMILES) # creates a column
for the intervals after the distance column
  mutate(TRPMILES = as.numeric(TRPMILES)) # makes sure the values are numeric
and not text in the column
data_TRIP_Final$TRPMILES <- ifelse(is.na(data_TRIP_Final$TRPMILES), 0,
data_TRIP_Final$TRPMILES) # makes sure there are no "NA" values left
(converted into "0")
for (s in 1:nrow(data_TRIP_Final)) {
  if (data_TRIP_Final$TRPMILES[s] >= distance_breaks_trip[1] &&
data_TRIP_Final$TRPMILES[s] < distance_breaks_trip[2]) {
    data_TRIP_Final$Distance_Intervals_Trip[s] <- "(0,0.5[" # then indicates
that intervals in the column with distance intervals
  } else if (data_TRIP_Final$TRPMILES[s] >= distance_breaks_trip[2] &&
data_TRIP_Final$TRPMILES[s] < distance_breaks_trip[3]) { # same for the
other intervals...
    data_TRIP_Final$Distance_Intervals_Trip[s] <- "(0.5,2["
  } else if (data_TRIP_Final$TRPMILES[s] >= distance_breaks_trip[3] &&
data_TRIP_Final$TRPMILES[s] < distance_breaks_trip[4]) {
    data_TRIP_Final$Distance_Intervals_Trip[s] <- "(2,5["
  } else if (data_TRIP_Final$TRPMILES[s] >= distance_breaks_trip[4] &&
data_TRIP_Final$TRPMILES[s] < distance_breaks_trip[5]) {
    data_TRIP_Final$Distance_Intervals_Trip[s] <- "(5,10["

```

```

    } else if (data_TRIP_Final$TRPMILES[s] >= distance_breaks_trip[5] &&
data_TRIP_Final$TRPMILES[s] < distance_breaks_trip[6]) {
      data_TRIP_Final$Distance_Intervals_Trip[s] <- "(10,25["
    } else if (data_TRIP_Final$TRPMILES[s] >= distance_breaks_trip[6] &&
data_TRIP_Final$TRPMILES[s] < distance_breaks_trip[7]) {
      data_TRIP_Final$Distance_Intervals_Trip[s] <- "(25,50["
    } else if (data_TRIP_Final$TRPMILES[s] >= distance_breaks_trip[7] &&
data_TRIP_Final$TRPMILES[s] < distance_breaks_trip[8]) {
      data_TRIP_Final$Distance_Intervals_Trip[s] <- "(50,100["
    } else if (data_TRIP_Final$TRPMILES[s] >= distance_breaks_trip[8] &&
data_TRIP_Final$TRPMILES[s] < distance_breaks_trip[9]) {
      data_TRIP_Final$Distance_Intervals_Trip[s] <- "(100,5000["
    }
  }
}

```

```

## >>> Travel time categories for each trip ("déplacement") :
library(dplyr) # charging libraries
travel_time_breaks_trip <- c(0,6, 11, 16, 21, 31, 46, 61,121,241,361,1441) ##
defining the limits for the intervals in minutes
travel_time.group_trip <- cut(as.numeric(data_TRIP_Final$TRVLCMIN), breaks =
travel_time_breaks_trip) # defining the intervals
data_TRIP_Final <- data_TRIP_Final %>% # python based function
  mutate(Travel_time_Intervals = NA, .after = TRVLCMIN) # creates a column for
the intervals after the time column
  mutate(TRVLCMIN = as.numeric(TRVLCMIN)) # makes sure the values are numeric
and not text in the column
data_TRIP_Final$TRVLCMIN <- ifelse(is.na(data_TRIP_Final$TRVLCMIN), 0,
data_TRIP_Final$TRVLCMIN) # makes sure there are no "NA" values left
(converted into "0")

```

```

for (t in 1:nrow(data_TRIP_Final)) {
  if (data_TRIP_Final$TRVLCMIN[t] >= travel_time_breaks_trip[1] &&
data_TRIP_Final$TRVLCMIN[t] < travel_time_breaks_trip[2]) {
    data_TRIP_Final$Travel_time_Intervals[t] <- "(0,5[" # then indicates that
intervals in the column with time intervals
  } else if (data_TRIP_Final$TRVLCMIN[t] >= travel_time_breaks_trip[2] &&
data_TRIP_Final$TRVLCMIN[t] < travel_time_breaks_trip[3]) { # same for the
other intervals...
    data_TRIP_Final$Travel_time_Intervals[t] <- "(5,10["
  } else if (data_TRIP_Final$TRVLCMIN[t] >= travel_time_breaks_trip[3] &&
data_TRIP_Final$TRVLCMIN[t] < travel_time_breaks_trip[4]) {
    data_TRIP_Final$Travel_time_Intervals[t] <- "(10,15["
  } else if (data_TRIP_Final$TRVLCMIN[t] >= travel_time_breaks_trip[4] &&
data_TRIP_Final$TRVLCMIN[t] < travel_time_breaks_trip[5]) {
    data_TRIP_Final$Travel_time_Intervals[t] <- "(15,20["
  } else if (data_TRIP_Final$TRVLCMIN[t] >= travel_time_breaks_trip[5] &&
data_TRIP_Final$TRVLCMIN[t] < travel_time_breaks_trip[6]) {
    data_TRIP_Final$Travel_time_Intervals[t] <- "(20,30["
  }
}

```

```

    } else if (data_TRIP_Final$TRVLCMIN[t] >= travel_time_breaks_trip[6] &&
data_TRIP_Final$TRVLCMIN[t] < travel_time_breaks_trip[7]) {
      data_TRIP_Final$Travel_time_Intervals[t] <- "(30,45["
    } else if (data_TRIP_Final$TRVLCMIN[t] >= travel_time_breaks_trip[7] &&
data_TRIP_Final$TRVLCMIN[t] < travel_time_breaks_trip[8]) {
      data_TRIP_Final$Travel_time_Intervals[t] <- "(45,60["
    } else if (data_TRIP_Final$TRVLCMIN[t] >= travel_time_breaks_trip[8] &&
data_TRIP_Final$TRVLCMIN[t] < travel_time_breaks_trip[9]) {
      data_TRIP_Final$Travel_time_Intervals[t] <- "(60,120["
    } else if (data_TRIP_Final$TRVLCMIN[t] >= travel_time_breaks_trip[9] &&
data_TRIP_Final$TRVLCMIN[t] < travel_time_breaks_trip[10]) {
      data_TRIP_Final$Travel_time_Intervals[t] <- "(120,240["
    } else if (data_TRIP_Final$TRVLCMIN[t] >= travel_time_breaks_trip[10] &&
data_TRIP_Final$TRVLCMIN[t] < travel_time_breaks_trip[11]) {
      data_TRIP_Final$Travel_time_Intervals[t] <- "(240,360["
    } else if (data_TRIP_Final$TRVLCMIN[t] >= travel_time_breaks_trip[11] &&
data_TRIP_Final$TRVLCMIN[t] < travel_time_breaks_trip[12]) {
      data_TRIP_Final$Travel_time_Intervals[t] <- "(360,1140]"
    }
  }
}

```

```

## >>> Persons on trip "Trajet" :
data_TRIP_Final <- data_TRIP_Final %>%
  mutate(On_Trip = NA, .after = NUMONTRP)

```

```

## >>> Multimodal per "trajet" (data_TRIP_Final) :
Superior_at_Final <- 1 #condition to only see the people with at least x
different transportation mode

```

```

library(dplyr)
data_TRIP_Final <- data_TRIP_Final %>%
  mutate(Multi_Trip = NA, .after = HOUSEID)

```

```

TYPE_COMBINED <- c(TYPE_PUBLIC_TRANSPORT, TYPE_DIVERS)
for (v in 1:(nrow(data_TRIP)-1)) { #max : "31073"
  if (data_TRIP_Final$TR_TYPE[v] != data_TRIP_Final$TR_TYPE[v+1] &&
data_TRIP_Final$PER_ID[v] == data_TRIP_Final$PER_ID[v+1]) {
    if (!data_TRIP_Final$TR_TYPE[v] %in% TYPE_NOT_CONSIDERED &&
!data_TRIP_Final$TR_TYPE[v+1] %in% TYPE_NOT_CONSIDERED){
      if (data_TRIP_Final$TR_TYPE[v] %in% TYPE_COMBINED ||
data_TRIP_Final$TR_TYPE[v+1] %in% TYPE_COMBINED) {
        data_TRIP_Final$Multi_Trip[v] <- 1
        data_TRIP_Final$Multi_Trip[v+1] <- 1
      }
    }
  }
}
data_TRIP_Final$Multi_Trip[is.na(data_TRIP_Final$Multi_Trip)] <- 0

```

```

## >>> % of trips multimodal :
nb_multi_Person <- sum(data_TRIP_Person$Multi == 1)
nb_total_Person <- length(data_TRIP_Person$PER_ID) #10591
Pourcent_Multi_Person <- 100*(nb_multi_Person/nb_total_Person)
print(sprintf("%.2f%%", Pourcent_Multi_Person))

nb_Multi_Trip <- sum(data_TRIP_Final$Multi_Trip == 1) # number of trips that
are multimodal
nb_total_Trip <- length(data_TRIP_Final$TDCASEID) # total number of trips
Pourcent_Multi_Trip <- 100*(nb_Multi_Trip/nb_total_Trip) # % of trips
multimodal in the total number of trips
print(sprintf("%.2f%%", Pourcent_Multi_Trip)) # shows results

library(knitr) # charging libraries
library(kableExtra)
Tbl_Multi_recap <- data.frame(c("Nombre de personne; multimodales", "Nombre
total de personnes", "% de personnes multimodales",
                             "Nombre de trajet multimodaux", "Nombre total
de trajet", "% de trajet multimodaux"),
                             c(round(nb_multi_Person,0),round(nb_total_Person
,0),round(Pourcent_Multi_Person,2),round(nb_Multi_Trip,0),
                             round(nb_total_Trip,0),round(Pourcent_Multi_Tr
ip,2))) # table do recap the previous results (ex: % of multimodals)
print(Tbl_Multi_recap) # shows results in a form of a table
kable(Tbl_Multi_recap, col.names = c("Description", "Value"), caption =
"Summary of Multimodal Persons and Trips") %>%
  kable_styling(bootstrap_options = c("striped", "hover", "condensed"),
full_width = F)

## >>> PRE-TESTS
## >>> Cramer's V (to adapt to the parameter) :
library(vcd) # charging libraries
cont_table <- table(data_TRIP_Person_V2$URBRUR, data_TRIP_Person_V2$Multi) #
Create a contingency table between the given parameter and 'MULTIMODAL'
variables
cramer_v_result <- assocstats(cont_table) # Perform the Cramer's V test
print(cramer_v_result) # shows results (Cramer results)
print(cont_table) # shows results (contingency table)

## >>> Wald test (to adapt to the parameter, based on simple regression on the
same parameter) :
library(car) # charging libraries
library(aod) # charging libraries
coef_est <- coef(Logistic_simple) # extract coefficient estimates and their
variance-covariance matrix
vcov_mat <- vcov(Logistic_simple) # co-valence matrix on a logistic regression
based of the selected variable

```

```

wald_statistic <- (coef_est['TRAVDAY'] / sqrt(vcov_mat['TRAVDAY',
'TRAVDAY']))^2 # wald test for 1 given variable
p_value <- 1 - pchisq(wald_statistic, df = 1) # calculates p-value
print(wald_statistic) # shows results (test results)
print(p_value) # shows results (p-value)

## >>> chi_square/PERSON test on a given variable (here "URBRUR") :
table_chi_square <- table(data_TRIP_Person_V2$URBRUR,
data_TRIP_Person_V2$Multi) # specific table based on the variable considered
in the dataset
chi_test_result <- chisq.test(table_chi_square) # chi-squared test applied to
a specific variable considered
print(chi_test_result) # shows results

## >>> ANOVA/GLOBAL LOGISTIC REGRESSION :
library(car) # charging libraries
data_TRIP_Person_V2$EDUC <- as.factor(data_TRIP_Person_V2$EDUC) # variable
transform into a factor for regression
data_TRIP_Person_V2$Age_Intervals <-
as.factor(data_TRIP_Person_V2$Age_Intervals) # variable transform into a
factor for regression
data_TRIP_Person_V2$R_SEX <- as.factor(data_TRIP_Person_V2$R_SEX) # variable
transform into a factor for regression
data_TRIP_Person_V2$LIF_CYC <- as.factor(data_TRIP_Person_V2$LIF_CYC) #
variable transform into a factor for regression
data_TRIP_Person_V2$HHFAMINC <- as.factor(data_TRIP_Person_V2$HHFAMINC) #
variable transform into a factor for regression
data_TRIP_Person_V2$R_RACE <- as.factor(data_TRIP_Person_V2$R_RACE) # variable
transform into a factor for regression
data_TRIP_Person_V2$WHYTO <- as.factor(data_TRIP_Person_V2$WHYTO) # variable
transform into a factor for regression
data_TRIP_Person_V2$PRMACT <- as.factor(data_TRIP_Person_V2$PRMACT) # variable
transform into a factor for regression
data_TRIP_Person_V2$DRIVER <- as.factor(data_TRIP_Person_V2$DRIVER) # variable
transform into a factor for regression
data_TRIP_Person_V2$TRAVDAY <- as.factor(data_TRIP_Person_V2$TRAVDAY) #
variable transform into a factor for regression
data_TRIP_Person_V2$Travel_time_Intervals <-
as.factor(data_TRIP_Person_V2$Travel_time_Intervals) # variable transform into
a factor for regression
data_TRIP_Person_V2$Distance_Intervals <-
as.factor(data_TRIP_Person_V2$Distance_Intervals) # variable transform into a
factor for regression
data_TRIP_Person_V2$URBANSIZE <- as.factor(data_TRIP_Person_V2$URBANSIZE) #
variable transform into a factor for regression
data_TRIP_Person_V2$URBRUR <- as.factor(data_TRIP_Person_V2$URBRUR) # variable
transform into a factor for regression
data_TRIP_Person_V2$TRPMILES <- as.factor(data_TRIP_Person_V2$TRPMILES) #
variable transform into a factor for regression

```

```

data_TRIP_Person_V2$TRVLCMIN <- as.factor(data_TRIP_Person_V2$TRVLCMIN) #
variable transform into a factor for regression
data_TRIP_Person_V2$DRIVERS <- as.factor(data_TRIP_Person_V2$DRIVERS) #
variable transform into a factor for regression
data_TRIP_Person_V2$HOUSEVEH <- as.factor(data_TRIP_Person_V2$HOUSEVEH) #
variable transform into a factor for regression
data_TRIP_Person_V2$ADULTS <- as.factor(data_TRIP_Person_V2$ADULTS) #
variable transform into a factor for regression

## >>> Problematic variables (almost continuous) -> time/distance/age :
data_TRIP_Person_V2$R_AGE <- as.factor(data_TRIP_Person_V2$R_AGE) # variable
transform into a factor for regression
data_TRIP_Person_V2$R_AGE <- as.numeric(data_TRIP_Person_V2$R_AGE) # The Age
stays a mono-category variable (don't explicit all the years)
data_TRIP_Person_V2$TRPMILES <- as.factor(data_TRIP_Person_V2$TRPMILES) #
variable transform into a factor for regression
data_TRIP_Person_V2$TRPMILES <- as.numeric(data_TRIP_Person_V2$TRPMILES) #
Distance stays a mono-category variable (don't explicit all the distances in
miles)
data_TRIP_Person_V2$TRVLCMIN <- as.factor(data_TRIP_Person_V2$TRVLCMIN) #
variable transform into a factor for regression
data_TRIP_Person_V2$TRVLCMIN <- as.numeric(data_TRIP_Person_V2$TRVLCMIN) #
Travel time stays a mono-category variable (don't explicit all the times in
minutes)

##>>> global logistic regression (all variables before tests and variables
eviction) :
Logistic_reg_Tot <- glm(Multi
~EDUC+Age_Intervals+R_SEX+LIF_CYC+HHFAMINC+R_RACE+WHYTO+PRMACT+DRIVER+TRAVDAY+
Travel_time_Intervals+
                        Distance_Intervals+URBANSIZE+URBRUR+DRIVERS+HOUSEVEH
+ADULTS+R_AGE+TRPMILES+TRVLCMIN, #all variables compared to "Multi" with a
glm model
                        data = data_TRIP_Person_V2, family = binomial(link =
"logit")) # glm on person tables (second version) following a binomial
distribution

## >>> global logistic regression (final variables after tests and variables
eviction) :
Logistic_reg_Tot <- glm(Multi ~
R_SEX+WHYTO+DRIVER+TRAVDAY+Travel_time_Intervals+Distance_Intervals+URBANSIZE+
HOUSEVEH+R_AGE, #all final variables compared to "Multi" with a glm model
                        data = data_TRIP_Person_V2, family = binomial(link =
"logit")) # glm on person tables (second version) following a binomial
distribution
summary(Logistic_reg_Tot) # details of the logistic regression

## >>> Logistic regression "Type III" (ANOVA) :
library(car) # changing libraries

```

```

Anova(Logistic_reg_Tot, type="III") # anova function and summary

## >>> Logistic regression almost continuous parameters
(R_AGE,TRPMILES,TRVLCMIN) :
Logistic_reg_continuous <- glm(Multi ~R_AGE+TRPMILES+TRVLCMIN, data =
data_TRIP_Person_V2, family = binomial(link = "logit"))
#summary(Logistic_reg_continuous)

## >>> Logistic regression 1 parameter (example with "Age_Intervals") :
data_TRIP_Person_V2$Age_Intervals<-
as.factor(data_TRIP_Person_V2$Age_Intervals  )
Logistic_simple <- glm(Multi ~ Age_Intervals , data = data_TRIP_Person_V2,
family = binomial(link = "logit"))
#summary(Logistic_simple)

##>>> GVIF test (multicollinearity) :
library(car) # charging libraries
vif_results <-vif(lm(Multi
~R_SEX+WHYTO+DRIVER+TRAVDAY+Travel_time_Intervals+Distance_Intervals+
URBANSIZE+HOUSEVEH+R_AGE+TRPMILES, data =
data_TRIP_Person_V2,type="GVIF")) # GVIF function on the final variables
selected
print(vif_results) # shows results

## >>> Hosmer-Lemeshow :
library(ResourceSelection) # charging libraries
hoslem_test_result <- hoslem.test(Logistic_reg_Tot$y,
fitted(Logistic_reg_Tot), g = 10) # on the final logistic model
print(hoslem_test_result) # shows results

##>>> ROC :
library(pROC)
predicted_probs <- predict(Logistic_reg_Tot, type = "response") # Predict
probabilities for the test
roc_obj <- roc(response = data_TRIP_Person_V2$Multi, predictor =
predicted_probs) # ROC on the final logistic model
c_statistic <- auc(roc_obj) # calculates AUC on the final logistic model
print(c_statistic) # shows results

## >>> MacFadden Rho Squared :
null_model <-glm(Multi ~ 1, data = data_TRIP_Person_V2, family = binomial(link
= "logit")) # calculates value without any variable
lokLik_full <- logLik(Logistic_reg_Tot) # log likelihood of the whole model
lokLik_null <- logLik(null_model) # log likelihood of the null model
mcfadden <- 1-(lokLik_full/lokLik_null) # actual test
print(mcfadden) # shows results

## >> Classification table :

```

```

predicted_probs <- predict(Logistic_reg_Tot, type = "response") # predict
probabilities
predicted_classes <- ifelse(predicted_probs > 0.5, 1, 0) # binarize
predictions at a threshold of 0.5
actual_outcomes <- data_TRIP_Person_V2$Multi # considered the "multi" column
(binary values : "1" or "2")
contingency_table <- table(Predicted = predicted_classes, Actual =
actual_outcomes) # create the contingency table
print(contingency_table)
## >>> apart from accuracy" everything is recalculated directly on the Excel !
library(caret) # charging libraries
confusionMatrix(as.factor(predicted_classes), as.factor(actual_outcomes)) #
calculate accuracy, sensitivity, etc. (complete results)
Misclassification_Rate =
(contingency_table[2,1]+contingency_table[1,2])/sum(contingency_table) #
calculated misclassification rate
Precision <- (contingency_table[2, 2]/sum(contingency_table[, 2]))*100
library(InformationValue) # charging libraries
Concordance(Logistic_reg_Tot$y,Logistic_reg_Tot$fitted.values) # calculate
concordance of the model

```

5.2 Sources

- Aoyong Li, & al. (2022). Comprehensive comparison of e-scooter sharing mobility : Evidence from 30 European cities. *Transportation Research Part D: Transport and Environment*, 105, 103229. <https://doi.org/10.1016/j.trd.2022.103229>
- Bai, Y., & Kattan, L. (2014). *Scopus—Document details—Modeling riders’ behavioral responses to real-time information at light rail transit stations | Signed in.* <https://doi.org/10.3141/2412-10>
- Bartlett, J. (2014a, février 8). R squared in logistic regression. *The Stats Geek*. <https://thestatsgeek.com/2014/02/08/r-squared-in-logistic-regression/>
- Bartlett, J. (2014b, mai 5). Area under the ROC curve – assessing discrimination in logistic regression. *The Stats Geek*. <https://thestatsgeek.com/2014/05/05/area-under-the-roc-curve-assessing-discrimination-in-logistic-regression/>
- Blumenberg, E., & Pierce, G. (2013). *Multimodal travel and the poor : Evidence from the 2009 National Household Travel Survey.* <https://doi.org/10.1179/1942787513Y.0000000009>
- Chieh-Hua Wen, & al. (2012). Latent class nested logit model for analyzing high-speed rail access mode choice. *Transportation Research Part E: Logistics and Transportation Review*, 48(2), 545-554. <https://doi.org/10.1016/j.tre.2011.09.002>
- Chorus, C. G., Arentze, T. A., & Timmermans, H. J. P. (2007). Information impact on quality of multimodal travel choices : Conceptualizations and empirical analyses. *Transportation*, 34(6), 625-645. <https://doi.org/10.1007/s11116-007-9120-1>
- Clauss, T., & Döppe, S. (2016). *Scopus - Document details - Why do urban travelers select multimodal travel options : A repertory grid analysis | Signed in.* <https://doi.org/10.1016/j.tra.2016.08.021>
- Çorbacioğlu, Ş. K., & Aksel, G. (2023). Receiver operating characteristic curve analysis in diagnostic accuracy studies : A guide to interpreting the area under the curve value. *Turkish Journal of Emergency Medicine*, 23(4), 195-198. https://doi.org/10.4103/tjem.tjem_182_23

- Couronné, R., Probst, P., & Boulesteix, A.-L. (2018). Random forest versus logistic regression : A large-scale benchmark experiment. *BMC Bioinformatics*, 19(1), 270.
<https://doi.org/10.1186/s12859-018-2264-5>
- datasciencebeginners. (2020, mai 27). *Binary Logistic Regression With R | R-bloggers*. <https://www.r-bloggers.com/2020/05/binary-logistic-regression-with-r/>
- Duthilleul, T., De Witte, L., Dubos, Robles, & Damien, M.-M. (1999). Pôles d'échanges et intermodalités, un enjeu pour la ville de demain, l'exemple de Lille-Flandres / Lille-Europe. *Hommes et Terres du Nord*, 4(1), 248-254. <https://doi.org/10.3406/htn.1999.2696>
- Ehab I. Diab, & Ahmed M. El-Geneidy. (2012). Understanding the impacts of a combination of service improvement strategies on bus running time and passenger's perception. *Transportation Research Part A: Policy and Practice*, 46(3), 614-625.
<https://doi.org/10.1016/j.tra.2011.11.013>
- Ehab I. Diab, & Ahmed M. El-Geneidy. (2014). *Transitory Optimism*. <https://doi.org/10.3141/2415-11>
- Ehab I. Diab, & al. (2015). *Bus Transit Service Reliability and Improvement Strategies : Integrating the Perspectives of Passengers and Transit Agencies in North America*.
<https://www.tandfonline.com/doi/epdf/10.1080/01441647.2015.1005034?src=getftr>
- Eric J.E. Molin, & al. (2010). Context Dependent Stated Choice Experiments : The Case of Train Egress Mode Choice. *Journal of Choice Modelling*, 3(3), 39-56. [https://doi.org/10.1016/S1755-5345\(13\)70013-7](https://doi.org/10.1016/S1755-5345(13)70013-7)
- Eva Heinen, & Kiron Chatterjee. (2015). *Scopus - Document details - The same mode again? An exploration of mode choice variability in Great Britain using the National Travel Survey | Signed in*. <https://doi.org/10.1016/j.tra.2015.05.015>
- Felix Schwinger, & al. (2022). *Scopus—Document details—Comparing Micromobility with Public Transportation Trips in a Data-Driven Spatio-Temporal Analysis | Signed in*.
<https://doi.org/10.3390/su14148247>

- Haochun, Y., & Yunyi, L. (2023). *Scopus—Document details—Examining the Connectivity between Urban Rail Transport and Regular Bus Transport | Signed in.*
<https://doi.org/10.3390/su15097644>
- Heinen, E., & Mattioli, G. (2019). Does a high level of multimodality mean less car use? An exploration of multimodality trends in England. *Transportation*, 46(4), 1093-1126.
<https://doi.org/10.1007/s11116-017-9810-2>
- Hensher, D. A., & Stopher, P. R. (Éds.). (2021). *Behavioural Travel Modelling*. Routledge.
<https://doi.org/10.4324/9781003156055>
- Hye Kyung Lee, & al. (2021). Identifying spatiotemporal transit deserts in Seoul, South Korea. *Journal of Transport Geography*, 95, 103145. <https://doi.org/10.1016/j.jtrangeo.2021.103145>
- Larousse, É. (2024, Consulté en). *Définitions : Multimodal - Dictionnaire de français Larousse.*
<https://www.larousse.fr/dictionnaires/francais/multimodal/10910241>
- Laura Alessandretti, & al. (2022). *Scopus—Document details—Multimodal urban mobility and multilayer transport networks | Signed in.* <https://doi.org/10.1177/23998083221108190>
- Li, J.-Y., Teng, J., & Wang, H. (2024). Measuring route diversity in spatial and spatial-temporal public transport networks. *Transport Policy*, 146, 42-58.
<https://doi.org/10.1016/j.tranpol.2023.11.008>
- Logistic Regression Essentials in R - Articles—STHDA.* (2018, mars 11).
<http://www.sthda.com/english/articles/36-classification-methods-essentials/151-logistic-regression-essentials-in-r/>
- Logit Regression | R Data Analysis Examples.* (2024, Consulté en).
<https://stats.oarc.ucla.edu/r/dae/logit-regression/>
- Marco Diana. (2010). From mode choice to modal diversion : A new behavioural paradigm and an application to the study of the demand for innovative transport services. *Technological Forecasting and Social Change*, 77(3), 429-441.
<https://doi.org/10.1016/j.techfore.2009.10.005>

Marco Diana. (2012). Measuring the satisfaction of multimodal travelers for local transit services in different urban contexts. *Transportation Research Part A: Policy and Practice*, 46(1), 1-11. <https://doi.org/10.1016/j.tra.2011.09.018>

Marco Diana, & al. (2009). Desire to change one's multimodality and its relationship to the use of different transport means. *Transportation Research Part F: Traffic Psychology and Behaviour*, 12(2), 107-119. <https://doi.org/10.1016/j.trf.2008.09.001>

Mode Choice to Modal Diversio, & al. (2017). *Modeling the Impact of Transit Fare Change on Passengers' Accessibility*. <https://doi.org/10.3141/2652-09>

Multinomial Logistic Regression | R Data Analysis Examples. (2024, Consulté en). <https://stats.oarc.ucla.edu/r/dae/multinomial-logistic-regression/>

National Household Travel Survey. (2024, Consulté en). <https://nhts.ornl.gov/>

Nattino, G., Pennell, M. L., & Lemeshow, S. (2020). Assessing the Goodness of Fit of Logistic Regression Models in Large Samples : A Modification of the Hosmer-Lemeshow Test. *Biometrics*, 76(2), 549-560. <https://doi.org/10.1111/biom.13249>

Nguyen Hoang-Tung, & al. (2021). Impacts of the introduction of bus rapid transit on travel behaviors of commuters in Hanoi, Vietnam : A quasi-experimental approach. *Case Studies on Transport Policy*, 9(1), 95-102. <https://doi.org/10.1016/j.cstp.2020.10.002>

Polydoropoulou, & Ben-AkivaA. (2001). *Combined Revealed and Stated Preference Nested Logit Access and Mode Choice Model for Multiple Mass Transit Technologies*. <https://doi.org/10.3141/1771-05>

Proc Logistic | SAS Annotated Output. (2024, Consulté en). <https://stats.oarc.ucla.edu/sas/output/proc-logistic/?fbclid=IwAR3qs90XrPtZmqAxMDgGinE4RuSSNv4VMZluelcWPnc8C5wuRkuxkfYpWbs>

Qu'est-ce que la régression logistique ? – Le modèle de régression logistique expliqué – AWS. (2024, Consulté en). Amazon Web Services, Inc. <https://aws.amazon.com/fr/what-is/logistic-regression/>

*R++, the Next Step*_C. Genolinia, J. Falcoub, et R. Tournierc (2010).pdf. (s. d.). Consulté 26 mai 2024, à l'adresse https://r2014-mtp.sciencesconf.org/38335/genoliniR_.pdf

R Advantages and Disadvantages—Javatpoint. (2024, Consulté en). www.javatpoint.com.
<https://www.javatpoint.com/r-advantages-and-disadvantages>

R: Cramer's V. (2024, Consulté en). https://search.r-project.org/CRAN/refmans/confintr/html/cramersv.html?fbclid=IwAR3KtL6OdBounA-LTq7HaG8p3MTUdgCPqX_-8jnWBy9ggBwOVwC94Ci4yic

Ralph Buehler, & al. (2007). *An examination of recent trends in multimodal travel behavior among American motorists*.
<https://www.tandfonline.com/doi/epdf/10.1080/15568318.2014.945672?src=getftr>

Resampling to Properly Handle Imbalanced Datasets in Machine Learning—Comet. (2024, Consulté en). <https://www.comet.com/site/blog/resampling-to-properly-handle-imbalanced-datasets-in-machine-learning/>

Riccardo Gallotti, & Marc Barthelemy. (2014). *Scopus—Document details—Anatomy and efficiency of urban multimodal mobility | Signed in*. <https://doi.org/10.1038/srep06911>

Saito, T., & Rehmsmeier, M. (2015). The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets. *PLOS ONE*, 10(3), e0118432.
<https://doi.org/10.1371/journal.pone.0118432>

Smart City Institute. (2021). *MAAS : UNE MEILLEURE ACCESSIBILITE AUX SERVICES EXISTANTS*.

Stephan Krygsman, & al. (2004). Multimodal public transport : An analysis of travel time elements and the interconnectivity ratio. *Transport Policy*, 11(3), 265-275.
<https://doi.org/10.1016/j.tranpol.2003.12.001>

Sungyop Kim, & al. (2007). Analysis of light rail rider travel behavior : Impacts of individual, built environment, and crime characteristics on transit access. *Transportation Research Part A: Policy and Practice*, 41(6), 511-522. <https://doi.org/10.1016/j.tra.2006.11.001>

- Thomas Klinger. (2017). *Scopus - Document details - Moving from monomodality to multimodality? Changes in mode choice of new residents | Signed in.*
<https://doi.org/10.1016/j.tra.2017.01.008>
- What's the Best R-Squared for Logistic Regression?* (2024, Consulté en). Sage Research Methods Community. <https://researchmethodscommunity.sagepub.com/blog/whats-the-best-r-squared-for-logistic-regression>
- William jen, & Kai-chieh hu. (2003). *Scopus - Document details - Application of perceived value model to identify factors affecting passengers' repurchase intentions on city bus : A case of the Taipei metropolitan area | Signed in.* <https://doi.org/10.1023/A:1023983627092>
- Wokje Abrahamse, & al. (2009). Factors influencing car use for commuting and the intention to reduce it : A question of self-interest or morality? *Transportation Research Part F: Traffic Psychology and Behaviour*, 12(4), 317-324. <https://doi.org/10.1016/j.trf.2009.04.004>
- Xavier Bach, & al. (2023). *Scopus - Document details - Spatial Inequalities in Access to Micromobility Services : An Analysis of Moped-Style Scooter Sharing Systems in Barcelona | Signed in.*
<https://doi.org/10.3390/su15032096>
- Yang Jiang, & al. (2012). Walk the line : Station context, corridor type and bus rapid transit walk access in Jinan, China. *Journal of Transport Geography*, 20(1), 1-14.
<https://doi.org/10.1016/j.jtrangeo.2011.09.007>
- Yifan Yue, & al. (2021). *Scopus—Document details—Evaluating the capacity coordination in the urban multimodal transport network | Signed in.* <https://doi.org/10.3390/app11178109>
- Yongsung Lee, Circella, G., Mokhtarian, P. L., & Guhathakurta, S. (2020). Are millennials more multimodal? A latent-class cluster analysis with attitudes and preferences among millennial and Generation X commuters in California. *Transportation*, 47(5), 2505-2528.
<https://doi.org/10.1007/s11116-019-10026-6>
- Yuji Sh. (2023). *Identifying the Spatiotemporal Metro-to-Bus Transfer Deserts in Shanghai, China.*
<https://doi.org/10.1061/JUPDDM.UPENG-4287>

Zhang, H., Shaheen, S. A., & Chen, X. (2014). Bicycle Evolution in China : From the 1900s to the Present. *International Journal of Sustainable Transportation*.

<https://www.tandfonline.com/doi/abs/10.1080/15568318.2012.699999>

Zihao An, & al. (2021). When you are born matters : An age-period-cohort analysis of multimodality.

Travel Behaviour and Society, 22, 129-145. <https://doi.org/10.1016/j.tbs.2020.09.002>

Zihao An, & al. (2022). Multimodal travel behaviour, attitudes, and cognitive dissonance.

Transportation Research Part F: Traffic Psychology and Behaviour, 91, 260-273.

<https://doi.org/10.1016/j.trf.2022.10.007>

Zimmerman, J. E., & Kramer, A. A. (2008). Outcome prediction in critical care : The Acute Physiology and Chronic Health Evaluation models. *Current Opinion in Critical Care*, 14(5), 491-497.

<https://doi.org/10.1097/MCC.0b013e32830864c0>