

# Facteurs de la fréquence d'utilisation des transports en commun urbains pour des trajets de personnes employées à Grenoble

Elizaveta Golovanova\*

Alexandra Lugova\*

## Résumé

Jusqu'à présent, la voiture reste la principale méthode de déplacement pour les trajets domicile-travail en France, comme dans la plupart des pays, ce qui entraîne des effets négatifs importants sur l'environnement, la santé et le bien-être de la société. Ce problème est particulièrement aigu dans les villes de taille moyenne, où malgré que les modes de transport doux deviennent de plus en plus populaires pour les courtes distances, mais la voiture personnelle prévaut encore fortement sur les transports en commun. Dans cette étude, un modèle de logit est construit sur les données des trajets des résidents de Grenoble pour identifier les facteurs qui influencent le choix du transport en commun comme un mode principal pour les déplacements des personnes employées. Selon les résultats, la présence d'une ou plusieurs voiture(s) dans le ménage reste l'un des principaux déclencheurs de la non-utilisation des transports en commun. Cela indique la nécessité de développer des politiques publiques visant à réduire l'attractivité de la possession et l'utilisation de la voiture privée et à encourager le choix des transports publics. En outre, selon notre modèle, la possession des permis de conduire, la disponibilité d'un abonnement de transport, la distance du trajet, la taille de ménage et l'âge d'un individu ont un impact statistiquement significatif (négatif sauf pour la disponibilité d'un abonnement, possession d'un ou plusieurs vélo(s) dans le ménage) sur l'utilisation des transports en commun.

---

\*BDA, MIASHS, Université Grenoble Alpes

# Table des matières

Introduction	3
Analyse de la littérature	3
Hypothèses et méthodologie	4
Analyse descriptive	4
Analyse statistique	7
Analyse économétrique: modèle de logit	10
Previsions et qualité du modèle	14
Limites	15
Conclusion: discussion	16
Bibliography	16

## Introduction

Il est largement reconnu que le système de transport contribue de manière significative aux problèmes environnementaux actuels tels que le réchauffement climatique, ainsi qu'aux problèmes de santé causés par les émissions et le manque d'activité physique. Dans les grandes et moyennes villes, le cadre de vie est négativement affecté par la circulation automobile. Le bruit, la pollution, les embouteillages et les risques d'accidents peuvent être réduits si les conducteurs actuels peuvent être encouragés à utiliser les transports en commun au lieu des voitures privées et si le besoin perçu de posséder des voitures peut être réduit. Pour que le système de transport contribue au développement durable, il faut rompre le lien entre croissance économique et croissance du transport motorisé. Étant donné que les densités de population élevées et les distances plus courtes offrent un plus grand potentiel pour des transports publics efficaces ainsi que pour la marche et le vélo, le potentiel de changement le plus élevé se trouve dans les zones urbaines.

Selon l'INSEE (Brutel et Pages (2021)), en 2017, en France, 74% des actifs en emploi qui déclarent se déplacer pour rejoindre leur lieu de travail utilisaient leur voiture, 16% prennent les transports en commun et 8% ont recours aux modes de transport doux (6% à la marche et 2% au vélo). Pour des distances inférieures à 5 kilomètres, la voiture représente encore 60% des déplacements domicile-travail, même si sa part diminue au profit des modes doux.

Au sein des villes moyennes, y compris à Grenoble, les habitants utilisent le plus les modes doux pour aller travailler, notamment le vélo. Quand même, le niveau de dépendance automobile reste aussi très élevé ce qui entraîne la nécessité d'ajouter les politiques des autorités locales afin d'inciter les habitants à utiliser plus fréquemment pas seulement les modes doux, mais aussi les transports en commun. Cet article vise à analyser les facteurs qui déterminent la fréquence de l'utilisation des transports en commun sur exemple des trajets effectués par les habitants employés du milieu urbain de Grenoble. Les résultats de cette étude pourraient être utilisés pour élaborer des politiques de réduction de la dépendance automobile en milieu urbain.

## Analyse de la littérature

La demande pour les différents modes de transport a été largement étudiée à l'aide de différents types de méthodes et de données, en focalisant sur contextes et facteurs différents. Ainsi, Holmgren et Ivehammar (2019) étudient la compétitivité des transports publics lorsque le vélo et la marche sont des options viables et constatent que la distance et les coûts financiers sont significativement importants dans le choix du mode de transport ainsi que l'accès à une voiture et les caractéristiques socio-démographiques des individus et des ménages. Pike et Lubell (2018) montrent que le choix des transports en commun des individus est influencé par le choix modal des autres personnes de leur réseau social ainsi que par les caractéristiques spécifiques des trajets domicile-travail. Ludfi, Melchior, et Achmad (2021) étudient le choix du mode de transport en tant que réponse émotionnelle à des incitations externes et montrent que l'expérience personnelle de voyage a un effet sur l'utilisation des transports publics par le changement des attitudes. Stiles et Smart (2021) montrent que la distance entre le lieu de travail et le domicile ainsi que les heures de pointe du trafic routier influencent significativement le choix de prendre ou non le trajet et le choix du mode de transport. Böcker, Amen, et Helbich (2017) prouvent que l'âge est un facteur important pour la fréquence d'utilisation des différents modes de transport.

Les études empiriques du choix modal du transport reposent sur deux principaux types de données. Dans les études dites de préférences révélées (Holmgren et Ivehammar (2019), Pike et Lubell (2018)), le comportement réel des personnes est observé et utilisé comme données. Au niveau individuel, les données peuvent être collectées à l'aide de questionnaires dans lesquels les schémas de déplacement des répondants sont documentés en détail. Le deuxième type de données qui peut être à la base pour les études sur le choix du mode de transport sont des données sur la préférence déclarée (Collins et Chambers (2005), Pike et Lubell (2018)). Ceux-ci sont obtenus en plaçant les répondants dans les enquêtes face à des choix hypothétiques. Par exemple, ils peuvent choisir entre différents modes de transport (voiture, bus, vélo ou marche) pour des situations où les prix de l'essence, les tarifs des bus, la durée du trajet et la météo varient. Ce type de données peut ensuite être analysé à l'aide des outils statistiques mentionnés ci-dessus. Les données sont ensuite analysées

à l'aide de modèles logit (Holmgren et Ivehammar (2019), Collins et Chambers (2005)) ou plus rarement probit, afin d'estimer les probabilités d'individus présentant des caractéristiques différentes de choisir un mode de déplacement particulier. Certains auteurs optent pour des méthodes d'estimation plus rares comme modèle "stimulus-organisme-réponse" (Ludfi, Melchior, et Achmad (2021)), théorie des réseaux et modèle d'inclusion résiduelle en deux étapes (Pike et Lubell (2018)) et autres.

## Hypothèses et méthodologie

Dans notre étude nous nous intéresserons à une question d'identification des déterminants de la fréquence d'utilisation des transports en commun urbains pour les trajets de personnes employées à Grenoble. Les hypothèses suivantes ont été élaborées sur la base des conclusions des auteurs de la littérature présentée ci-dessus, de la disponibilité et de la qualité des données :

$H_1$ : La fréquence de l'utilisation des transports en commun diminue quand la distance de trajet est très grande à cause de la préférence vers la voiture (Brutel et Pages (2021), Holmgren et Ivehammar (2019), Pike et Lubell (2018), Stiles et Smart (2021)).

$H_2$ : La fréquence de l'utilisation des transports en commun diminue lorsqu'un individu a un accès aux autres moyens de transport (possession d'une voiture ou d'un vélo, possession des permis de conduire) et diminue même plus si plusieurs voitures / vélos sont disponibles (Holmgren et Ivehammar (2019), Ludfi, Melchior, et Achmad (2021), Stiles et Smart (2021)).

$H_3$ : La fréquence de l'utilisation des transports en commun diminue quand la taille d'un ménage augmente (Holmgren et Ivehammar (2019)).

$H_4$ : Si un individu voyage souvent aux heures de pointe, lorsque le trafic sur les routes est intense, la fréquence de l'utilisation des transports en commun augmente (Stiles et Smart (2021)).

Pour tester les hypothèses listées ci-dessus, tout d'abord, la statistique descriptive ainsi qu'analyse statistique vont être menés. En premier lieu, les caractéristiques de l'échantillon seront étudiées et examinées pour la présence des valeurs aberrantes. Ensuite, les coefficients de corrélation de Pearson entre la variable cible et les variables explicatives seront calculés et testés pour la significativité statistique. Par ailleurs, les tests statistiques seront appliqués, selon le type de données, pour tester la significativité de la différence des caractéristiques des individus utilisant régulièrement le transport en commun par rapport à ceux qui l'utilisent rarement ou n'utilisent pas du tout: le test du  $\chi^2$  de Pearson pour les variables binaires et le test de Wilcoxon pour les variables discrets et continues. Enfin, l'analyse de la variance ANOVA sera effectuée pour vérifier encore une fois la différence statistique de ces deux groupes d'individus.

Le modèle de logit, étant le plus couramment utilisé dans la littérature sur les choix de mode de transport (ex. Holmgren et Ivehammar (2019), Collins et Chambers (2005)), sera ensuite entraîné sur les données. Les variables seront ensuite choisies pour optimiser le modèle et assurer la qualité la plus haute possible à l'aide d'un algorithme éliminant pas à pas les caractéristiques non-significatives et déterminant le modèle avec le critère d'information d'Akaike (AIC) le plus faible. Les coefficients et les "odds ratios" du modèle optimisé seront interprétés. Finalement, les conclusions sur la qualité du modèle seront faites pour valider les résultats en divisant l'échantillon en parties d'entraînement et de test, faisant des prédictions sur les données de test et évaluant la précision du modèle en les comparant avec les données réelles.

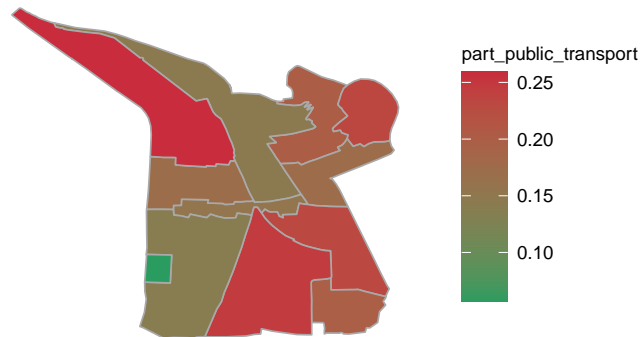
## Analyse descriptive

Pour avoir une idée générale de la base de données que nous allons utiliser, regardons la fréquence d'utilisation des transports en commun en général à Grenoble. Nos données montrent les déplacements de personnes pour un jour en 2009. Nous avons filtré les personnes qui utilisent les transports en commun (bus urbain ou tram) plus de deux fois par semaine et divisé la valeur obtenue par chaque unité territoriale. A titre de comparaison,

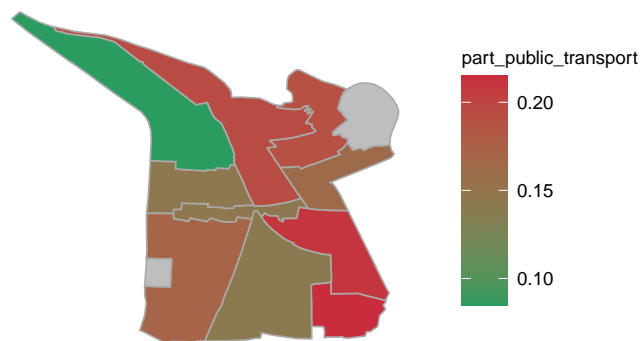
nous montrons également une carte de l'intensité d'utilisation des transports en commun par le lieu de travail des répondants (Figure 1). Les valeurs pour la deuxième carte sont calculé dans la même façon.

Figure 1 - De haut en bas: Carte de la fréquence d'utilisation des transports en commun en général à Grenoble et carte de l'intensité d'utilisation des transports en commun par le lieu de travail des répondants.

Intensité de l'utilisation des transports en commun à Grenoble dans chaque TIR



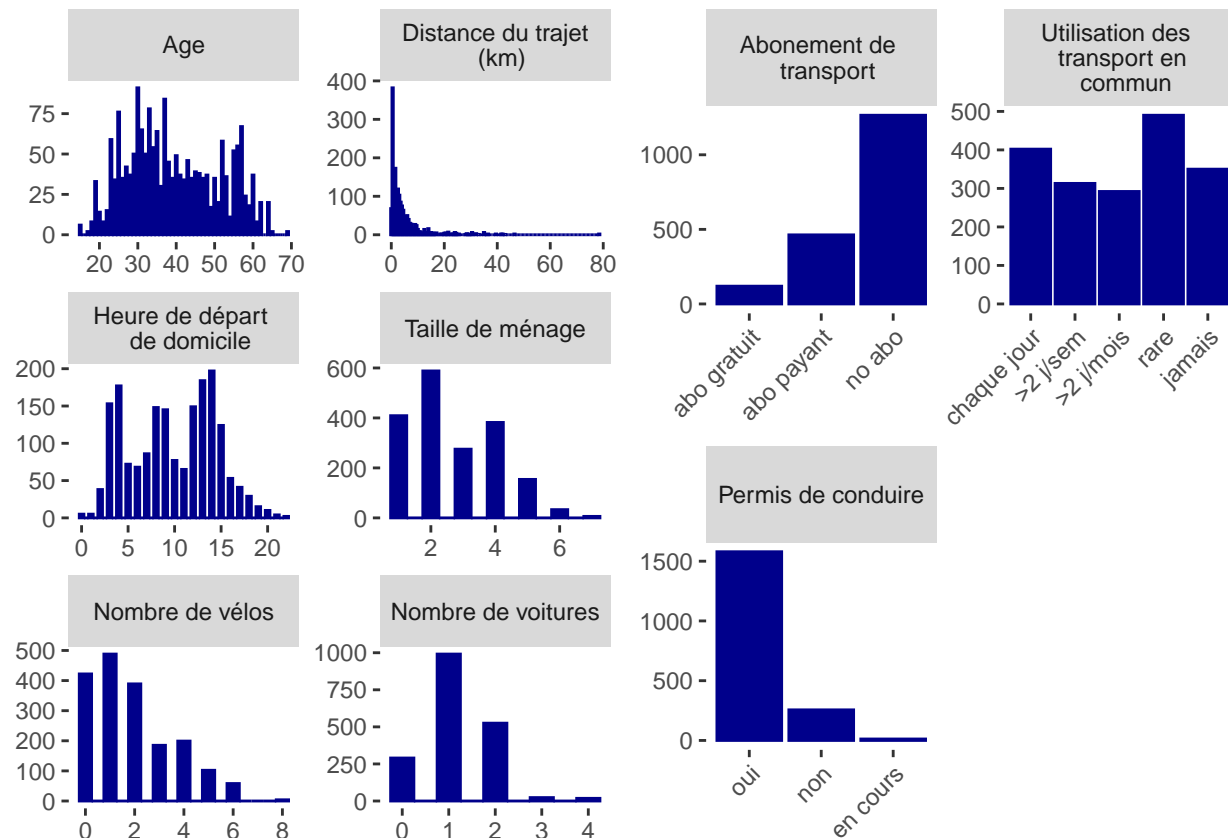
Intensité de l'utilisation des transports en commun à Grenoble par zone de travail dans chaque TIR



Les têtes de liste des quartiers par utilisation des transports en commun sont Centre ville, Berriat Nord, Ile Verte, Reyniès-Bayard Village Olympique, Malherbe Capuche. Quartiers de Grenoble avec au moins 20% des trajets en transport commun par zone de travail sont Malherbe Capuche et Villeneuve de Grenoble.

Pour l'étude, nous avons sélectionné uniquement les répondants qui ont signalé qu'ils sont employés dans le temps plein ou partiel, ainsi que des stagiaires et alternants. Après avoir fait l'analyse de la littérature, on a choisi 9 variables pertinentes pour notre étude : fréquence d'utilisation des transport en commun, âge de la personne, possession d'un permis de conduire, possession d'un abonnement de transport, taille de ménage, nombre de vélos et voitures dans le ménage, mesure de la distance en km et heure de départ de domicile. La figure 2 montre des graphiques de ces variables.

Figure 2 - Distributions des variables sélectionnées



L'âge de la plupart de répondants est environ 25-35 ans. Dans la plupart des cas, le trajet ne dépasse pas 5 kilomètres, mais il existe également des distances plus longues (jusqu'à 80 kilomètres). Le plus souvent, les enquêtés quittent leur domicile à 3h00-4h00 et à 8h00-9h00 du matin ou de 12h00 à 14h00. En moyenne, un ménage comprend 2 personnes et possède 1 voiture et 0-2 vélo(s). La majorité des répondants interrogés n'ont pas d'abonnement aux transports en commun et ne les utilisent que rarement. De plus, une partie importante possède un permis de conduire. Pour l'analyse suivant, nous avons transformé certaines des variables, comme indiqué dans le tableau 1. Il est à noter, qu'à 3h00-4h00 du matin il n'y a pas de trafic, alors on ne va pas considérer ces valeurs comme les heures de pointe.

Tableau 1 - Variables transformés pour l'analyse.

Nom	Explication	Variable
frequcu	Fréquence de l'utilisation du transport en commun	1 - plus que 2 fois par semaine 0 - plus rarement
permis	Possession des permis de conduire	1 - oui 0 - non
ABO_TC	Possession d'un abonnement de transport	1 - oui 0 - non
nb_pers	Nombre de personnes dans le ménage	
VP_dispo	Nombre de voitures dans le ménage	
NB_velo	Nombre de vélos dans le ménage	
D13	Distance domicile-travail en km	
heuredep	Heure de départ de domicile pour le lieu de travail	1 - départ aux heures de pointe (8-9, 12-14)

Nom	Explication	Variable
		0 - autres
age	Age de la personne	

Examinons les statistiques descriptives des variables transformées (Tableau 2). Au total, il y a 1977 observations dans notre échantillon, et certaines variables contiennent de valeurs manquantes, notamment, **freqtcu**, **D13** et **heuredep**. L'âge d'un répondant moyen est 40 ans. 38% de l'échantillon se rendent au travail en transports en commun plus que 2 jours par semaine. 86% ont un permis de conduire et 31% possèdent un abonnement aux transports en commun. En moyenne, les répondants parcourent 4,34 kilomètres pour se déplacer. Cependant, il convient de noter que l'écart-type de cette variable est grand et fait plus que la moyenne (6.65). 44% des déplacements se font aux heures de pointe. En moyenne, les ménages sont composés de trois personnes, ils possèdent 1 voiture et 1-2 vélos. Après qu'on a supprimé les valeurs manquants, il reste 1847 observations, ou 12% d'échantillon initiale.

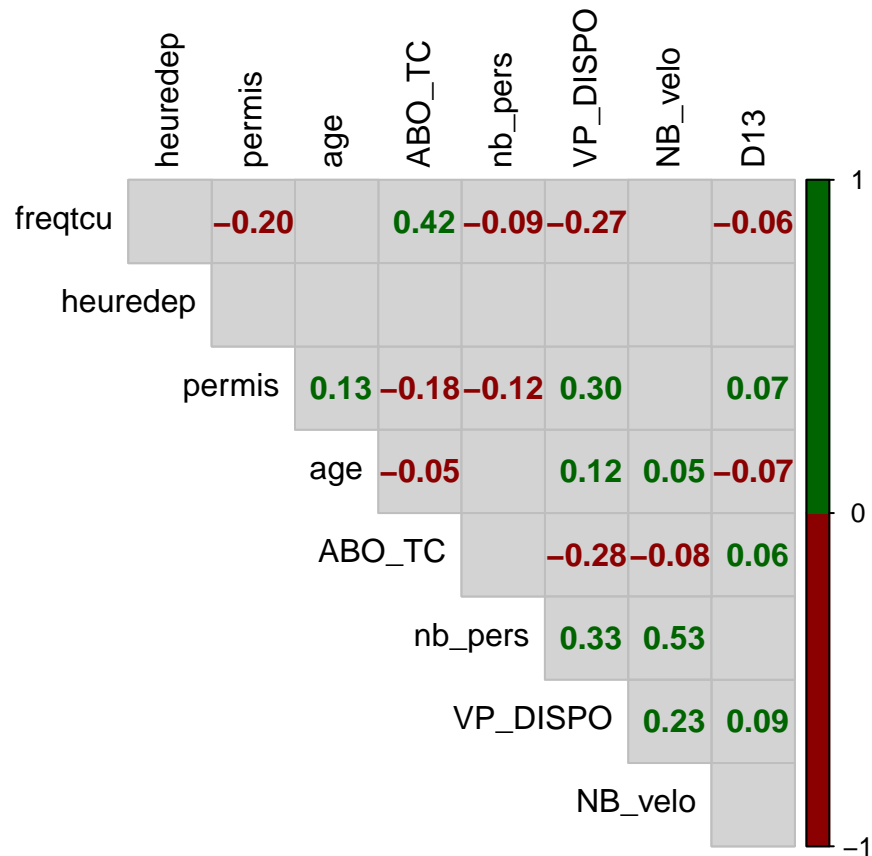
Tableau 2 - Statistiques descriptives de la base de données.

	vars	n	mean	sd	min	max	range	se
freqtcu	1	1974	0.38	0.48	0	1.0	1.0	0.011
permis	2	1977	0.86	0.35	0	1.0	1.0	0.008
ABO_TC	3	1977	0.31	0.46	0	1.0	1.0	0.010
age	4	1977	39.48	11.87	15	69.0	54.0	0.267
nb_pers	5	1977	2.70	1.36	1	7.0	6.0	0.031
VP_DISPO	6	1977	1.20	0.75	0	4.0	4.0	0.017
NB_velo	7	1977	1.91	1.66	0	8.0	8.0	0.037
D13	8	1905	4.39	6.65	0	78.3	78.3	0.152
heuredep	9	1954	0.44	0.50	0	1.0	1.0	0.011

## Analyse statistique

Afin de valider le choix des variables pour l'estimation et la prédiction de la fréquence d'utilisation des transports en commun à Grenoble, nous analysons dans un premier temps les coefficients de corrélation de Pearson. Le graphique ci-dessous représente la matrice de corrélation pour l'ensemble de variables. La matrice affiche uniquement les coefficients qui sont statistiquement significatifs au niveau de 5%. On constate bien qu'il y a une corrélation significative entre la plupart des variables. Pour notre variable cible la corrélation avec **NB\_velo**, **age** et **heuredep** s'est avéré pas significative. Quand même, il convient de noter que la corrélation montre une relation linéaire sans tenir compte de l'influence d'autres variables, bien que lorsqu'elles sont ajoutées au modèle, certaines variables peuvent montrer une signification conjointe et certaines relations peuvent ne pas être linéaires. Alors, la fréquence d'utilisation de transport en commun a une corrélation négative avec la taille de ménage, possession d'un permis de conduire et une voiture, ainsi qu'avec la distance de trajet. Il y a une corrélation positive et modérée de la variable ciblé seulement avec la possession d'un abonnement aux transports en commun.

Figure 3 - Matrice de corrélation



Pour valider la rationalité des hypothèses posées, il faut également vérifier que les groupes des utilisateurs et non-utilisateurs des transports en commun à Grenoble sont statistiquement différents selon les caractéristiques choisies. Le tableau suivant présente les résultats d'une série de tests statiques effectués pour chaque variable selon le type de données. On peut constater que le groupe d'individus employés qui utilisent fréquemment les transports publics est statiquement différent (au seuil de 1%) de ceux qui les utilisent rarement ou pas du tout selon toutes les caractéristiques sauf l'heure de départ, l'âge et la distance du trajet.

Tableau 3 - Comparaison de groupes des utilisateurs des transports en commun selon la fréquence d'utilisation.

label	variable	freqtcu		test
		0	1	
ABO_TC	0	950 (75.10%)	315 (24.90%)	p value: <0.0001 (Pearson's Chi-squared test)
	1	182 (31.27%)	400 (68.73%)	
permis	0	102 (38.06%)	166 (61.94%)	p value: <0.0001 (Pearson's Chi-squared test)
	1	1030 (65.23%)	549 (34.77%)	
VP_DISPO	Min / Max	0 / 4.0	0 / 3.0	p value: <0.0001 (Wilcoxon rank sum test)
	Med [IQR]	1.0 [1.0;2.0]	1.0 [0;1.0]	
	Mean (std)	1.3 (0.7)	0.9 (0.7)	
	N (NA)	1132 (0)	715 (0)	



label	variable	freqtcu		test	
		0	1		
heuredep	0	611 (59.67%)	413 (40.33%)	p value: 0.1107 (Pearson's Chi-squared test)	
	1	521 (63.30%)	302 (36.70%)		
D13	Min / Max	0 / 43.1	0.1 / 78.3	p value: 0.1189 (Wilcoxon rank sum test)	
	Med [IQR]	2.4 [0.8;5.8]	2.3 [0.9;4.5]		
	Mean (std)	4.7 (6.6)	3.9 (6.6)		
	N (NA)	1132 (0)	715 (0)		
nb_pers	Min / Max	1.0 / 6.0	1.0 / 7.0	p value: <0.0001 (Wilcoxon rank sum test)	
	Med [IQR]	3.0 [2.0;4.0]	2.0 [1.0;4.0]		
	Mean (std)	2.8 (1.3)	2.5 (1.4)		
	N (NA)	1132 (0)	715 (0)		
age	Min / Max	15.0 / 69.0	19.0 / 65.0	p value: 0.3446 (Wilcoxon rank sum test)	
	Med [IQR]	37.0 [30.0;47.0]	39.0 [30.0;52.0]		
	Mean (std)	39.0 (11.3)	39.9 (12.9)		
	N (NA)	1132 (0)	715 (0)		
NB_velo	Min / Max	0 / 8.0	0 / 6.0	p value: 0.0263 (Wilcoxon rank sum test)	
	Med [IQR]	2.0 [1.0;3.0]	1.0 [1.0;3.0]		
	Mean (std)	2.0 (1.7)	1.8 (1.7)		
	N (NA)	1132 (0)	715 (0)		

L'analyse de la variance ANOVA permet également de vérifier que les moyennes des groupes des utilisateurs et non-utilisateurs des transports en commun proviennent des populations différentes. Les groupes sont significativement différents au seuil de 1% sauf l'heure de départ (pas de différence significative entre les groupes) et taille de ménage (significative au seuil de 10%). Pour la première version du modèle économétrique on va conserver quand même ces variables. En générale, les résultats des tests réalisés nous permet de confirmer la raisonabilité de notre analyse et de passer à l'étape de la recherche de relations causales afin de déterminer les facteurs et leurs influençant le choix du mode de transport pour les trajets des personnes employées, ainsi que la direction et le force des effets.

Tableau 4 - Résultats du test ANOVA

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
ABO_TC	1	76.57	76.57	411.99	0.00
permis	1	6.49	6.49	34.93	0.00
VP_DISPO	1	7.14	7.14	38.44	0.00
heuredep	1	0.36	0.36	1.94	0.16
D13	1	1.71	1.71	9.21	0.00
nb_pers	1	0.73	0.73	3.93	0.05

	DfSum	SqMean	SqF	valuePr(>F)	
NB_velo	1	0.98	0.98	5.26	0.02
age	1	2.65	2.65	14.24	0.00
Residuals	1838	341.59	0.19	NA	NA

## Analyse économétrique: modèle de logit

Un modèle logit est construit et entraîné sur les données pour estimer les effets causals des variables explicatives choisies (la possession d'une ou plusieurs voiture(s) ou d'un ou plusieurs vélo(s), la possession des permis de conduire, la possession d'un abonnement de transport, le nombre de personnes dans le ménage, l'heure de départ de domicile et la distance du trajet, l'âge) sur la fréquence d'utilisation des transports publics. On peut aussi supposer que l'âge a une relation plus complexe avec la variable dépendante. Nous ajouterons un carré de cette variable pour contrôler cela. Le modèle peut être formalisé par l'équation suivante:

$$p(freqtcu_i = 1) = \frac{1}{1 + f(\Lambda)}$$

où

$$f(\Lambda) = f(permis, ABO\_TC, nb\_pers, age, age^2, VP\_dispo, NB\_velo, D13, heureddep)$$

Afin de maximiser la qualité de modèle, il est possible de reconsidérer le choix des variables. En utilisant un algorithme qui compare, étape par étape, les modèles avec différents ensembles de variables sur la base du critère d'Akaike, nous déterminons l'ensemble optimal de variables qui fournit l'AIC le plus bas et qui devrait être laissé dans le modèle final.

Tableau 5 - Variables sélectionnées sur la base du test d'Akaike

	LR	Chisq	Df	Pr(>Chisq)
permis	14.784106	1	0.0001205	
nb_pers	4.807045	1	0.0283436	
age	8.247790	1	0.0040802	
I(age^2)	11.568270	1	0.0006709	
VP_DISPO	27.663220	1	0.0000001	
NB_velo	8.004947	1	0.0046650	
ABO_TC	241.055887	1	0.0000000	
heureddep	2.640293	1	0.1041843	
D13	6.229299	1	0.0125653	

En résultat, on supprime de notre modèle initiale la variable **heureddep**, indiquant le fait de partir de la maison en heurs de pointe. Dans le tableau 6, la première colonne montre les résultats de l'estimation des coefficients du modèle complet avec toutes les variables incluses en utilisant la méthode du maximum de vraisemblance. Les estimations des coefficients du modèle final sont présentées dans la deuxième colonne de ce tableau. Nous avons également calculé le coefficient de détermination de McFadden (Pseudo- $R^2$ ). Il est défini comme  $1 - LL_{mod}/LL_0$  où  $LL_{mod}$  est la valeur du log de vraisemblance pour le modèle ajusté et  $LL_0$  est le log de vraisemblance pour le modèle nul qui n'inclut qu'une interception comme prédicteur.

Le pseudo- $R^2$  de McFadden allant de 0.2 à 0.4 indique un très bon ajustement du modèle. Alors, pour le modèle complet, ainsi que pour le modèle final il est égal au 0.18. Nous avons également essayé de supprimer la variable d'âge au carré, ce qui a réduit le coefficient de détermination à 0.17.

Tableau 6 - Résultats de la régression logistique

	freqtcu	
permis	-0.625*** (0.163)	-0.638*** (0.163)
nb_pers	-0.113** (0.052)	-0.113** (0.052)
age	-0.097*** (0.034)	-0.097*** (0.034)
I(age^2)	0.001*** (0.0004)	0.001*** (0.0004)
VP_DISPO	-0.477*** (0.093)	-0.476*** (0.092)
NB_velo	0.114*** (0.040)	0.113*** (0.040)
ABO_TC	1.761*** (0.118)	1.756*** (0.118)
heuredep	-0.178 (0.110)	
D13	-0.022** (0.009)	-0.021** (0.009)
Constant	1.740*** (0.638)	1.677*** (0.637)
Observations	1,847	1,847
Log Likelihood	-1,013.381	-1,014.701
Akaike Inf. Crit.	2,046.761	2,047.401

Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Tous les coefficients du modèle final sont statistiquement significatifs au seuil de 5%. En interprétant les coefficients, on peut observer les effets suivantes:

## effets négatifs

- la probabilité qu'un individu employé utilise les transports en commun à Grenoble diminue s'il a les permis de conduire (validation de  $H_2$ );
- la probabilité qu'un individu employé utilise les transports en commun à Grenoble diminue si le nombre de voiture dans sa possession augmente (validation de  $H_2$ );
- la probabilité qu'un individu employé utilise les transports en commun à Grenoble diminue avec l'augmentation de la taille de ménage (validation de  $H_3$ );
- la probabilité qu'un individu employé utilise les transports en commun à Grenoble diminue avec l'augmentation de la distance du trajet (validation de  $H_1$ ).

## effets positifs

- la probabilité qu'un individu employé utilise les transports en commun à Grenoble augmente s'il a un abonnement de transport;
- la probabilité qu'un individu employé utilise les transports en commun à Grenoble augmente si le nombre de vélos dans sa possession augmente (contradiction avec  $H_2$ ).

L'analyse des ratios de probabilité ("odds ratios") nous permet d'aller plus loin dans l'interprétation des effets estimés par le modèle.

Tableau 7 - Résultats pour les odds-ratios

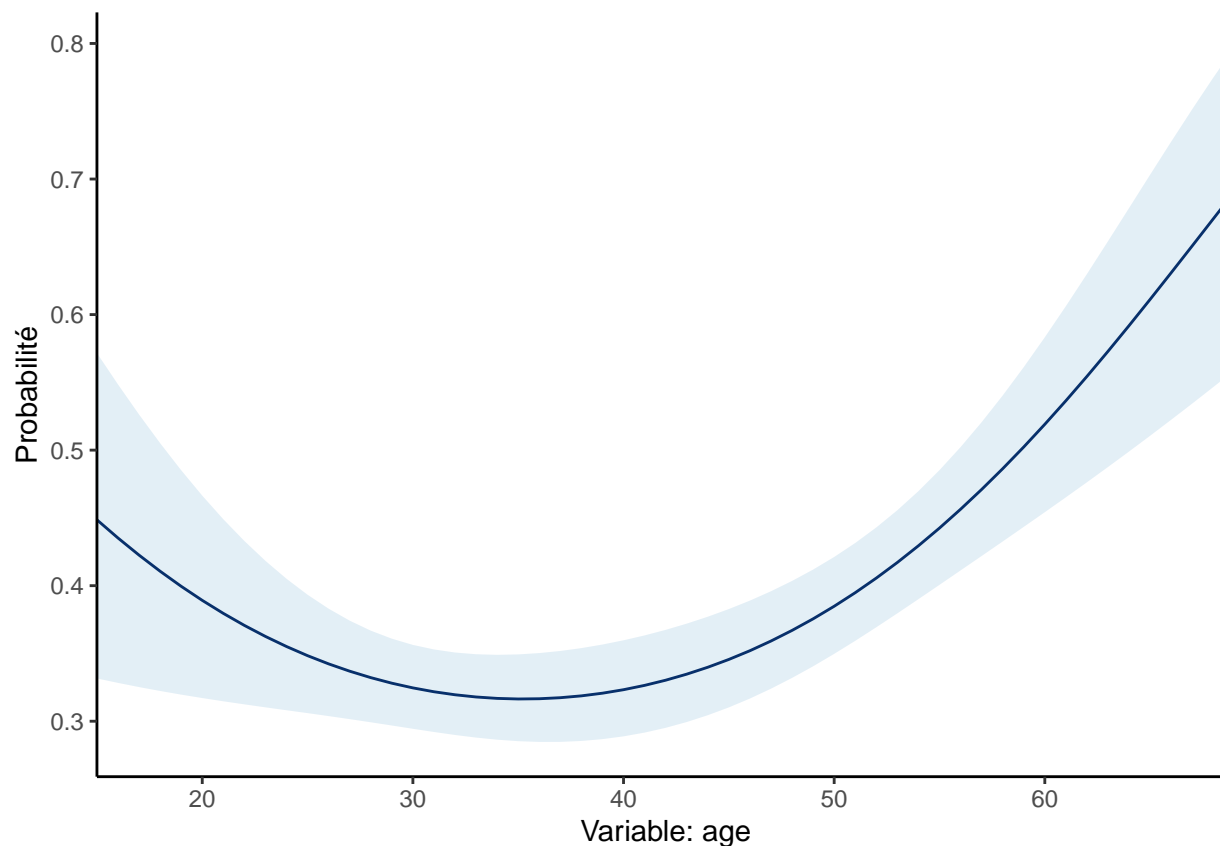
explanatory OR(multivarié)	
permis	0.52 (0.38-0.71, p<0.001)
VP_DISPO	0.65 (0.55-0.78, p<0.001)
NB_velo	1.10 (1.02-1.19, p=0.018)
ABO_TC	5.75 (4.58-7.24, p<0.001)
D13	0.97 (0.96-0.99, p=0.005)
nb_pers	0.87 (0.79-0.96, p=0.006)

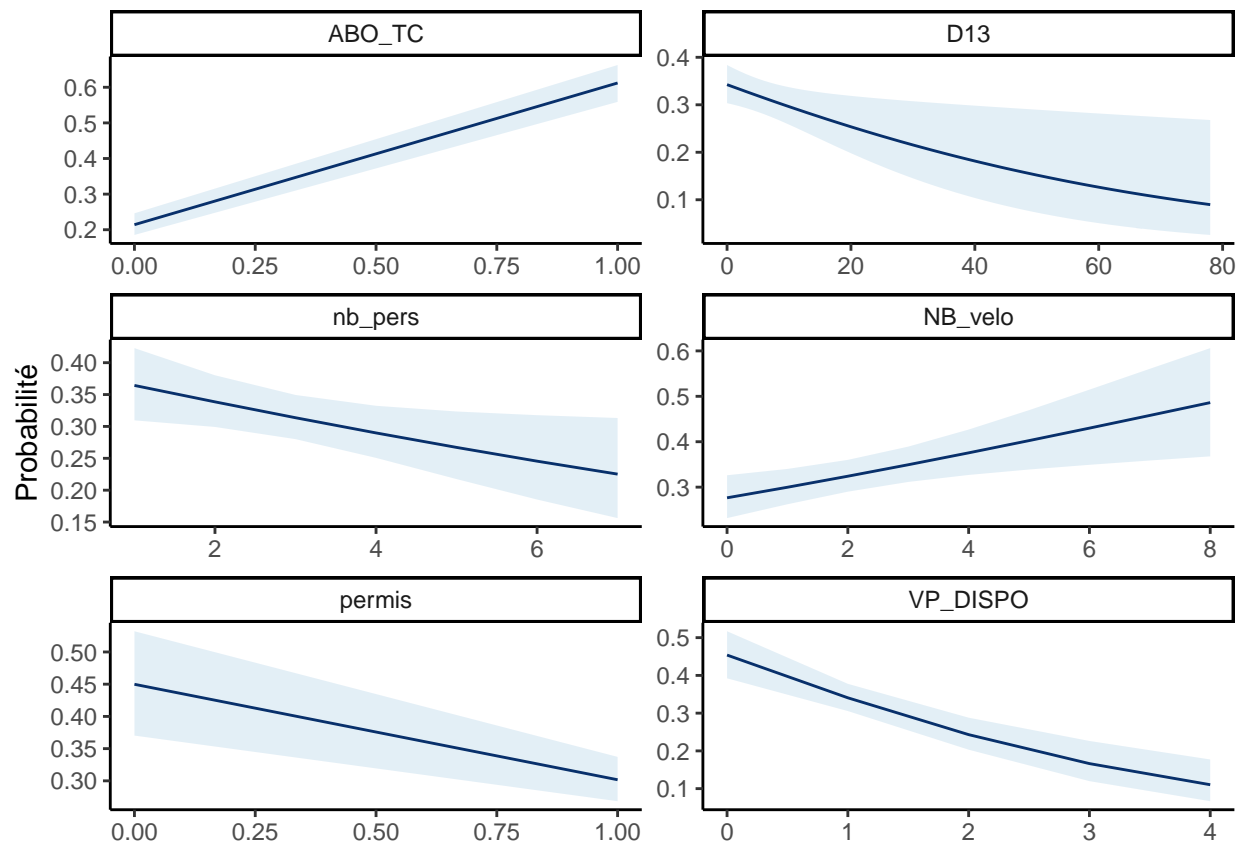
De tableau 7 on peut conclure que:

- En gardant toutes les autres variables constantes, la possession par un individu employé des permis de conduire augmente son ratio de probabilité d'utiliser les transports en commun de 0.52 fois;
- En gardant toutes les autres variables constantes, une voiture supplémentaire dans le ménage augmente le ratio de probabilité pour un individu employé d'utiliser les transports en commun de 0.65 fois;
- En gardant toutes les autres variables constantes, un vélo supplémentaire dans le ménage augmente le ratio de probabilité pour un individu employé d'utiliser les transports en commun de 1.1 fois;
- En gardant toutes les autres variables constantes, la possession par un individu employé d'un abonnement de transport augmente son ratio de probabilité d'utiliser les transports en commun d'environ 5.34 fois;
- En gardant toutes les autres variables constantes, lorsque la taille de ménage augmente d'une personne, il est 0.87 fois plus susceptible d'être dans une catégorie des utilisateurs de transport en commun;
- En gardant toutes les autres variables constantes, lorsque distance du trajet augmente d'un kilomètre, il est 0.97 fois plus susceptible d'être dans une catégorie des utilisateurs de transport en commun.

Pour mieux comprendre l'interprétation de odds ratio, on a calculé les effets marginaux pour les variables dans le modèle final. L'effet marginal est généralement considéré comme un point, plus souvent pour une observation moyenne, car en utilisant le modèle logit, nous devons prendre en compte une relation exponentielle et non linéaire. Cependant, pour plus de clarté et pour une meilleure compréhension de la dynamique d'évolution des effets marginaux, nous avons décidé de tracer des graphiques montrant l'évolution de la probabilité d'utiliser les transports en commun pour une observation moyenne sur toute la plage des valeurs disponibles. Par exemple, pour estimer l'effet de la possession de voiture, il faut calculer son effet marginal sur la variable **fregtcu** :  $\frac{d\hat{p}(fregtcu_i=1)}{dVP\_dispo} = \frac{\exp^{-(1.677-0.477*VP\_dispo)}}{(1+\exp^{-(1.677-0.477*VP\_dispo)})^2} * (-0.477)$ . Alors, pour la **VP\_dispo** = 1 l'effet marginal est de -0.085. Cela signifie qu'une augmentation de la quantité de voitures dans le ménage de 1 à 2 diminuera la probabilité d'utilisation de transport en commun de 0.085. On a fait une illustration pour les effets marginaux de tous les variables dans le modèle (Figure 4).

Figure 4 - Les effets marginaux pour les variables de modèle final.





Dans notre échantillon, il s'est avéré que l'hypothèse  $H_4$  est rejetée, car la variable indiquant les heures de points n'était pas significative. Donc, la fréquence de l'utilisation des transports en commun ne change pas si un individu employé voyage souvent aux heures de pointe, lorsque le trafic sur les routes est intense ( $H_4$ ).

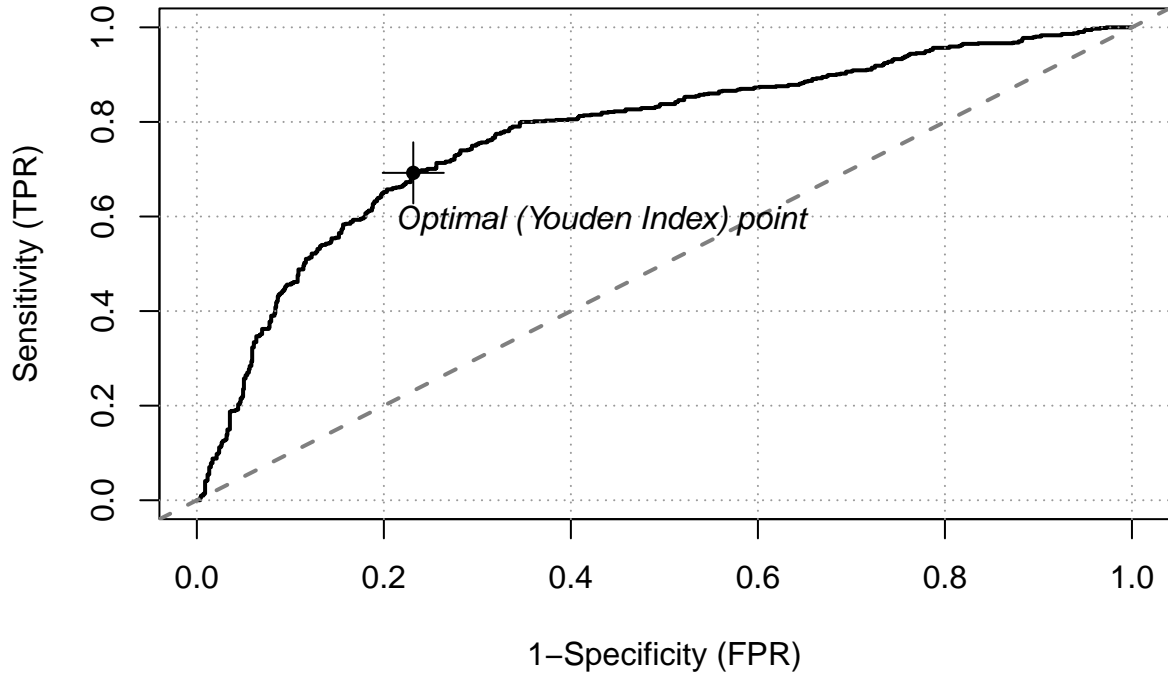
## Previsions et qualité du modèle

Regardons la qualité du modèle final résultant, en s'appuyant sur la courbe ROC. La courbe ROC (Receiver Operating Characteristic) représente la sensibilité en fonction de  $1 - \text{spécificité}$  pour toutes les valeurs seuils possibles du marqueur étudié. La sensibilité est la capacité du test à bien détecter les personnes qui utilisent le transport en commun et la spécificité est la capacité du test à bien détecter les non-utilisateurs du transport en commun. L'aire sous la courbe ROC (ou Area Under the Curve, AUC) peut être interprétée comme la probabilité que, parmi deux sujets choisis au hasard, un individu employé qui utilise souvent le transport en commun et un individu qui le utilise rarement ou jamais, la valeur du marqueur soit plus élevée pour l'utilisateur de transport en commun. Par conséquent, une AUC de 0,5 (50%) indique que le marqueur est non-informatif. Une augmentation de l'AUC indique une amélioration des capacités discriminatoires, avec un maximum de 1,0 (100%). Figure 4 montre la courbe ROC empirique, ainsi que l'indice de Youden (ou statistique J de Youden) qui montre le point de cutoff optimal. J de Youden est défini comme :

$$J = \max(\text{sensitivity} + \text{specificity} - 1)$$

L'AUC sous la courbe ROC est égal à 77.5%. Le cutoff optimale calculé de la fonction pour logit est de 0.3677226.

Figure 5 - La courbe ROC.



En divisant les données en un échantillon de test (30%) et un échantillon d'apprentissage (70%) et en testant le modèle, nous prouvons la qualité assez élevée: le modèle prédit la fréquence d'utilisation des transports en commun avec une précision de 73%.

## Limites

L'échantillon que nous utilisons est une des limites majeures de notre étude. La base de données ne contient pas tous les déplacements des individus pour 2009 et on ne connaît pas bien la représentativité de cet échantillon. Cela pourrait potentiellement affecter les variables indiquant les heures de pointe et la distance du trajet. De plus, il serait intéressant de regarder les données actuelles, car pendant 12 ans, il pourrait y avoir beaucoup de changements liés aux infrastructures et aux besoins des personnes. Par exemple, le nombre de personnes employées dans le travail à distance a évidemment augmenté au cours des 3 dernières années.

Il serait également possible d'examiner les autres variables comme les dépenses financières de transport pour tester l'hypothèse de savoir si une augmentation des dépenses financières d'essence, par exemple, affecte l'utilisation des transports publics. Ce serait intéressant à tester, surtout dans le contexte de la situation géopolitique actuelle et de la hausse des prix du carburant en France. Selon Holmgren et Ivehammar (2019), Pike et Lubell (2018), Collins et Chambers (2005), la fréquence de l'utilisation des transports en commun diminue quand les coûts financiers liés à cela sont élevés.

Nous sommes également contraints de constater le problème d'endogénéité potentielle dans notre modèle, puisqu'il n'est pas possible d'étudier l'influence mutuelle des variables, par exemple, la présence d'un abonnement aux transports en commun (`ABO_TC`) et la fréquence d'utilisation des transports en commun (`frequcu`): soit une personne prend les transports en commun car elle a déjà acheté un abonnement, soit elle a acheté un abonnement car elle utilise souvent les transports en commun. Ainsi, les résultats obtenus doivent être interprétés en prenant en compte cette remarque.

## Conclusion: discussion

Cette étude a utilisé les données des caractéristiques de répondants, leur ménages et leurs déplacements afin d'étudier les facteurs de la fréquence de choix des transports en commun pour des individus employés à Grenoble. La possession d'un abonnement de transport en commun ou des permis de conduire, la possession d'un ou plusieurs vélo(s) et voitures, la distance du trajet, la taille de ménage et l'âge d'un individu se sont révélés comme des facteurs important du choix des transports en commun comme mode principal.

Un résultat important robuste de l'analyse est que le fait d'avoir accès à une voiture, ou plus précisément le nombre de voitures dans le ménage, réduit considérablement la probabilité qu'une personne choisisse tout autre mode de transport. Ce résultat a des implications politiques importantes. Parallèlement à l'augmentation des revenus et du volume de la circulation, la possession d'une voiture a augmenté en France comme dans la plupart des autres pays. Étant donné que les propriétaires de voitures sont difficiles à influencer pour qu'ils utilisent des modes de transport plus respectueux de l'environnement, il est particulièrement important de rompre ce lien pour l'évolution de l'impact environnemental futur du secteur des transports.

La rupture du lien entre revenus et possession d'une voiture peut évidemment être obtenue en réduisant l'attractivité de la possession ou de l'utilisation d'une voiture particulière, en augmentant l'attractivité des autres modes ou par une combinaison de ces changements. Le statut et d'autres motifs sociaux peuvent également affecter le choix du mode et le choix de posséder une voiture. La combinaison de l'amélioration de la fonctionnalité des transports publics et des tarifs réduits avec d'autres mesures visant spécifiquement à réduire l'attractivité de l'utilisation de la voiture est nécessaire pour obtenir des résultats significatifs. Ces mesures pourraient inclure l'extension des infrastructures de transport public, l'augmentation des frais de stationnement ainsi que la restriction de l'accès en voiture à certaines parties de la ville. Une approche encore plus large de la question consisterait à mettre l'accent sur la mobilité respectueuse de l'environnement dans la planification urbaine.

## Bibliography

- Böcker, L., P. van Amen, et M. Helbich. 2017. « Elderly travel frequencies and transport mode choices in Greater Rotterdam, the Netherlands ». *Transportation* 44: 831-52. <https://doi.org/10.1007/s11116-016-9680-z>.
- Brutel, C., et J. Pages. 2021. « La voiture reste majoritaire pour les déplacements domicile-travail, même pour de courtes distances ». *Insee Première*, n 1835: 1-4.
- Collins, C. M., et S. M. Chambers. 2005. « Psychological and Situational Influences on Commuter-Transport-Mode Choice ». *Environment and Behavior* 37 (5): 640-61. <https://doi.org/10.1177/0013916504265440>.
- Holmgren, J., et P. Ivehammar. 2019. « Mode choice in home-to-work travel in mid-size towns: The competitiveness of public transport when bicycling and walking are viable options ». *Transportation Research Procedia* 48 (septembre): 1635-43. <https://doi.org/10.1016/j.trpro.2020.08.204>.
- Ludfi, D., B. Melchior and Melchior, et W. Achmad. 2021. « How Employees Choose their Commuting Transport Mode: Analysis Using the Stimulus-Organism-Response Model ». *Journal of Advanced Transportation* 2021 (novembre): 1-16. <https://doi.org/10.1155/2021/5555488>.
- Pike, S., et M. Lubell. 2018. « The conditional effects of social influence in transportation mode choice ». *Research in Transportation Economics* 68: 2-10.
- Stiles, J., et M. J. Smart. 2021. « Working at home and elsewhere: daily work location, telework, and travel among United States knowledge workers ». *Transportation* 48: 2461-91. <https://doi.org/10.1007/s11116-020-10136-6>.