

Facteurs influençants la prise des transports en commun pour l'agglomération grenobloise

RACHIDI Mustapha & SAUNIER Florent & SAADALLAH Malek

Janvier 2023

#J'ai tenté l'image

```
knitr::include_graphics("Image_BUS_TRAM.jpg")
```



Introduction

Ce projet se base sur des données récoltées en 2010 dans la région Grenobloise. L'étude a pour but de déterminer les facteurs influençant la prise des transports en commun. Pour cela nous nous sommes pris comme limites : le réseau Mtag qui comprend les bus qualifiés de "ville" (Nous n'avons pas pris en compte les bus régionaux comme par exemple le bus Grenoble - Chamrousse) et le réseau du tramway dont les lignes depuis 2010 ont été augmentées.

Articles de la littérature

Familiarisation avec la base de données

La base de données contient 30 702 lignes et 116 colonnes ce qui correspond à nos variables, on peut la qualifier de base de données "moyenne" mais qui saura nous occuper. Concernant le nombre de valeurs manquantes, toutes variables confondues nous avons 971 658 valeurs manquantes soit 27.3% de notre base de données. De plus, 0% des lignes ont toutes leurs valeurs et c'est 21% des colonnes qui n'ont pas de valeurs manquantes. Il peut être intéressant de voir où sont les valeurs manquantes.

L'échantillon comporte 5189 personnes

Visualisation valeurs manquantes titre à changer peut être

En annexe, quelques graphiques permettant de visualiser quelles variables ont le plus de valeurs manquantes. Ces graphiques nous permettront d'adopter un regard critique sur les variables que nous choisirons par la suite. Cependant, on peut établir quelques critères avec r : ration de valeurs manquantes dans la colonne.

Bon : $r \leq 5\%$ Moyen : $5\% < r \leq 20\%$ Mauvais $20\% < r \leq 45\%$ Très mauvais : $r > 45\%$

Plusieurs variables ont entre 80% 99% de valeurs manquantes

Variables du projet

Frecqtcu : Variable d'intérêt (Y) catégorielle qui indique la fréquence d'utilisation des transports en communs chez une personne.

Elle prend les valeurs :

- 1 : Utilisation des transports en commun tous les jours
- 2 : Utilisation des transports en commun au moins deux fois par semaine
- 3 : Utilisation des transports en commun au moins deux fois par mois
- 4 : Utilisation des transports en commun très rare
- 5 : Utilisation des transports en commun inexistante

Nous avons décidé de construire frecqtcu de manière à ce qu'elle prenne la valeur 0 ou 1

```
DB_projet_full<-DB_projet_full%>%mutate(frecqtcu=ifelse(frecqtcu<=3,1,0))
DB_projet_full$frecqtcu<-factor(DB_projet_full$frecqtcu)
```

Pour toutes les personnes qui prennent les transports de manière : régulière/tous les jours, au moins deux fois par semaine et au moins deux fois par mois se sont vues attribuées la valeur 1 car le "au moins" présage une prise des transports en communs plus élevée.

Tailmng : Variable qui indique le nombre de personnes composant le ménage.

```
DB_projet_full<-rename(DB_projet_full,"tailmng"="NO_PERS")
```

On change simplement le nom de la variable “NO_PERS” qui indique le nombre de personne dans le ménage
Permis : Variable indiquant si la personne effectuant le trajet possède le permis ou pas.

```
DB_projet_full<-DB_projet_full%>%mutate(permis=ifelse(any(permis==1 | permis==3),"YES","NO"))
DB_projet_full$permis<-factor(DB_projet_full$permis)
```

Car_ownership : Variable indiquant si la personne effectuant le trajet possède une voiture

```
DB_projet_full<-DB_projet_full%>%mutate(car_ownership=ifelse(DB_projet_full$VP_DISPO>0 & (DB_projet_full$GENRE=="M" | DB_projet_full$POSSE=="O"), "YES", "NO"))
DB_projet_full$car_ownership<-factor(DB_projet_full$car_ownership)
```

Cette variable dépend de trois variables qui sont VP_dispo qui doit être strictement supérieur à 0, puis GENRE (type de véhicule utilisé), nous avons exclu les campings cars car notre sujet se prête au milieu urbain et de POSSE (Est ce que la voiture appartient à la personne). Nous nous sommes contentés de prendre exclusivement les véhicules possédés par la personne.

Création de la nouvelle base de données

Variables complémentaires

Grâce aux variables précédentes et aux articles que l’on a trouvé dans la littérature, nous allons construire notre base de données pour notre modèle.

Nous exploiterons un ensemble de caractéristiques socio-économiques puis certaines variables liées au “confort” du trajet.

Restriction géographique Définissons ce que l’on entend par “transports en communs”.

Pour notre étude nous nous concentrons sur les transports en communs de la société MTag, c’est à dire les tram et bus du réseau.

Notre délimitation géographique sera simplement les terminaux des différentes lignes de tram/bus confondues.

Par la suite, quand on parlera de transports en communs, on se réfère à la définition au dessus.

```
Vec_zone<-c(101,102,103,104,105,106,107,108,109,110,111,112,113,114,117,118,135,119,120,136,121,122,123)
```

Toutes les zones répertoriées dans le vecteur “Vec_zone” ont au moins un arrêt du réseau Mtag.

```
New_DB<-dplyr::filter(New_DB,tir %in% Vec_zone) #on garde que les zones où il y a des transports en communs
```

Restriction sur l’âge

Il est nécessaire de préciser que les mineurs se déplacent majoritairement via les transports en communs car ils n’ont tout simplement pas le choix...

Pour ne pas être biaisé, il est judicieux de filtrer les mineurs de notre base de données ainsi que les personnes âgées de plus de 80ans.

```
New_DB<-dplyr::filter(New_DB,age>=18 & age<=80)
```

Notre nouvelle base de données comprend maintenant 10 879 observations et 22 variables ## Analyse Univariée

Analyse Univariée

```
summary(New_DB$freqtcu)
```

```
##      0      1 NA's  
## 5823 5028    28
```

```
counts<-table(New_DB$freqtcu)  
(counts/nrow(New_DB))*100
```

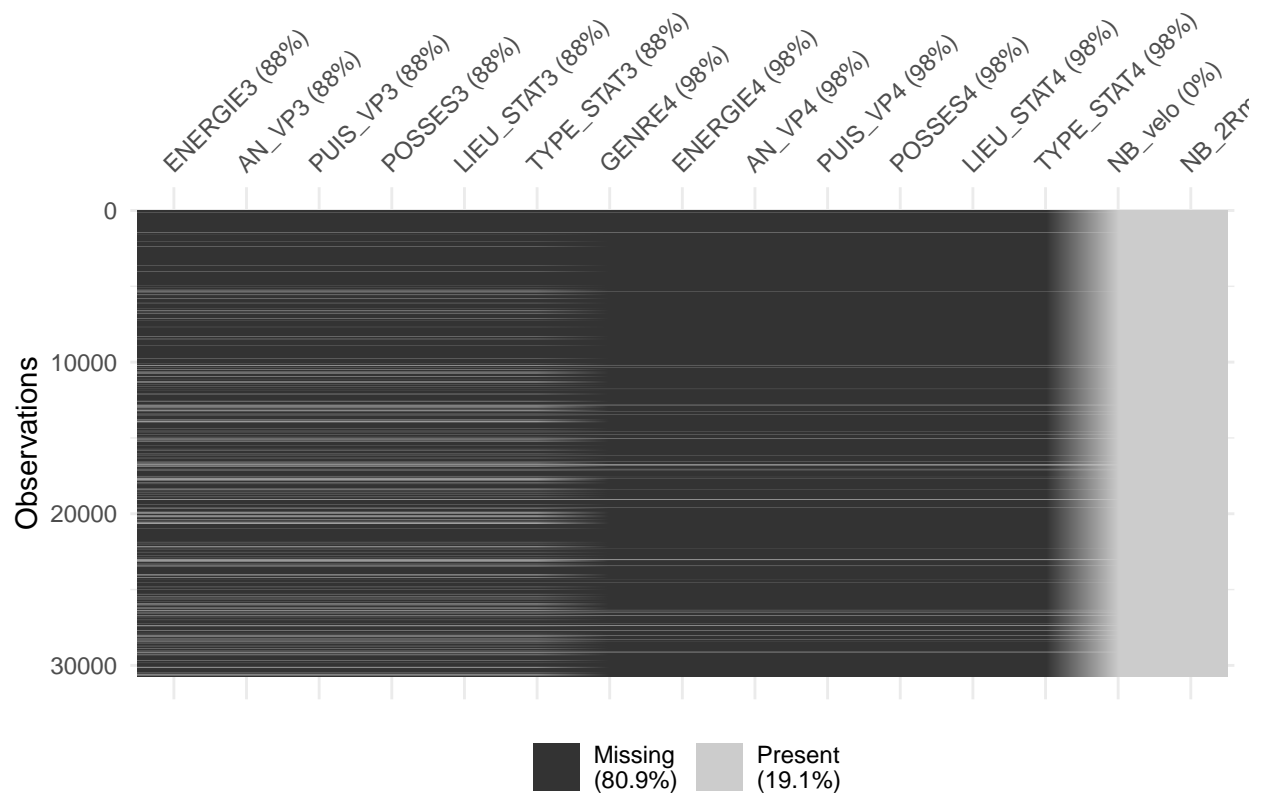
```
##  
##      0      1  
## 53.52514 46.21748
```

#percentage

Dans notre base de données, il y a 46% des gens qui prennent les transports en communs de manière plus ou moins régulière. **Analyse Bivariée**

Annexes

```
data_2<-DB_projet_full[,c(30:44)]
vis_miss(
  data_2,
  cluster = FALSE,
  sort_miss = FALSE,
  show_perc = TRUE,
  show_perc_col = TRUE,
  large_data_size = 9e+06,
  warn_large_data = TRUE
)
```



Listes variables à plus de 80% de valeurs manquantes

-motoracc -situveil -STAT_TRAV -TYPE_STAT4 -LIEU_STAT4 -POSSES4 -PUIS_VP4 -AN_VP4 -
ENERGIE4 -GENRE4 -TYPE_STAT3 -LIEU_STAT3 -POSSES3 -PUIS_VP3 -AN_VP3 -ENERGIE3 -
motdeacc -nbarret -abonpage