

Projet Econométrie 3

Variables qualitatives et modèles de choix discret

I. Joly

29 octobre 2020

Table of Content

Descriptif du TP Econométrie 3	2
Enoncé	2
Descriptif du fichier de données <i>allgre.PB_V2.Rdata</i>	3
Liste des fichiers joints:	3
Le thème et les données	3
Liste des sujets proposés	3
Management des Données	5
Partie 1 - Preprocessing	5
Chargement de la base de données et exploration	5
Chargement de la base	5
Partie 2 - Data processing	5
Description du contenu de la base de données	6
Description des variables et labels	7
Structure de la base de données	9
Définition de nouvelles variables	10
Personne Immobile	10
Redéfinition des numéro de zone	10
Distance parcourue moyenne par zone	10
Définition de quelques variables	11
Constitution des 3 bases	12
Constitution de la base <i>Déplacements</i>	12
Constitution de la base <i>Individus</i>	13
Constitution de la base <i>Ménages</i>	14
Partie 3 - Sauvegarde des Data	14
Partie - Représentation SIG	15
Import des fichiers shp avec coordonnées compatibles	17
Contrôle des shapefiles	17
Carte et calcul des indicateurs par zone: CARTE NBD BTT	19
Jointure newESRI et data	19
Carte colorée	19
Carte avec labels	20
Carte complète des durées quotidiennes de déplacement moyennes par zone	21

Descriptif du TP Econométrie 3

Enoncé

- Ce projet est à faire en **binôme**.
- Tous les binômes travaillent sur le **même jeu de données**, avec des attentes identiques, mais avec un angle différent.
- Plus précisément, chaque binôme a une **question à traiter** (dans la liste des questions proposée en fin de sujet - l'affectation des questions aux groupes sera déterminée aléatoirement)
- Chaque énoncé précise trois variables (X, Y, Z) d'intérêt particulier pour cette question. Vous **incluez systématiquement ces trois variables** à vos analyses.

Vous devez rendre un **rapport de 10 pages maximum** (plus annexes) abordant les points suivants, que vous mettrez en perspective avec la question à traiter :

1. Présenter et décrire le jeu de données que vous utilisez (sans privilégier a priori une variable plutôt qu'une autre) : caractéristiques remarquables, propriétés des différentes variables, structures sous-jacentes, création de variables, factorisation, etc.
2. Étude **univariée** des trois variables
3. Etudier la variable X , **variable à expliquer** (facteurs permettant de connaître X ? modèles de prédiction ?)
4. Étudier la variable Z et sa contribution à l'explication de X
5. Étudier la variable Y et sa contribution à l'explication de X

Pour chacun des points, précisez le cas échéant si vous avez travaillé sur un sous-ensemble d'observations et/ou si vous avez supprimé ou créé des variables.

De plus, n'hésitez pas à collaborer avec vos camarades travaillant sur d'autres variables, les problématiques étant liées et les résultats des uns pouvant être exploités aussi par d'autres.

Le corps du rapport doit être une synthèse de vos analyses, où vous présentez les éléments que vous estimez les plus importants. Des éléments concrets (valeurs, tableaux, figures) doivent étayer vos affirmations, mais ne donnez pas tous les détails de vos analyses dans le corps du rapport (les annexes sont faites pour cela). Par exemple, si vous voulez étudier les corrélations de variables, n'indiquez que les plus significatives dans le rapport, et renvoyez aux annexes pour la matrice complète des corrélations.

Le rapport devra être réalisé au format Rmarkdown (ou KnitR) avec tous les éléments permettant de **reproduire** votre analyse, et rendu par email (iragael.joly@grenoble-inp.fr ou ijoly.research@gmail.com) le **10/01/2021** à 23h59 dernier délai.

Il est **impératif d'assurer la reproductibilité** de votre travail, tant sur la manipulation des données que sur la production du rapport. Veillez donc à indiquer à l'utilisateur les packages à installer les fichiers intermédiaires à exécuter, etc.

Descriptif du fichier de données *allgre.PB_V2.Rdata*

Liste des fichiers joints:

- Base de données **allgre.PB_V2.RData**
- Dictionnaire de variables **DICO_VAR_EMD_GRE2010.xls**
- Fichier **EnonceTP_Eco3_2017.rmd** de cet énoncé. Ce fichier Rmd crée les fichiers utilisables pour le TP. Il est vivement conseillé de l'utiliser pour initier les manipulations des données.
- Fichiers SIG: Tous les fichiers nécessaires pour l'application SIG sont dans le dossier : **FOND_MAP**
- Les codes SIG sont décrits: **Decoupage_simplifie.xlsx**

Le thème et les données

Les données utilisées sont issues d'une extraction d'une enquête sur les déplacements des grenoblois en 2010. Ces données sont disponibles au format **environnement R** : *allgre.PB_V2.Rdata*

Les données décrivent les déplacements réalisés par les membres de ménages grenoblois en 2010. Elles renseignent des données relatives aux caractéristiques individuelles et du ménage (structure du ménage, motorisation, sexe, âge, statut professionnel, etc.) et aux pratiques de mobilité (nombres de déplacements, horaires des déplacements, localisations des déplacements, modes utilisés, motifs des déplacements, etc.).

La structure de la base de données est décrite dans les sections suivantes.

Liste des sujets proposés

1. Décrire les durées de déplacements quotidiennes (variable BTT¹) et comparer les deux groupes suivants : 1) les personnes se déplaçant moins de 2h par jour 2) les personnes se déplaçant plus de 2h par jour.

Variable X : **btt**

Variable Y : **jourdepl**

Variable Z : **permis**

2. Décrire le choix 'voiture' du mode de déplacement des individus et identifier les déterminants de ce choix parmi les caractéristiques individuelles ou du ménage et des modes de transport.

Variable X : **Voiture** (variable à créer selon le `mode_depl_ag='VP'` ou non)

Variable Y : **permis**

Variable Z : **age**

3. Proposer une caractérisation de la population de personnes immobiles et identifier les déterminants de la mobilité

Variable X : **immobil** (variable à créer selon le `nbd= 0` ou `nbd>0`)

Variable Y : **age**

Variable Z : **VP_DISPO**

4. Quels sont les déterminants de la motorisation des ménages ?

Variable X : **motorisation** (variable à créer selon `VP_DISPO = 0` ou `VP_DISPO >0`)

Variable Y : **taillemng**

Variable Z : **Permis_mng** : au moins 1 permis dans le ménage

5. Etudier le parc automobile de Grenoble. Quels sont les déterminants du choix d'une motorisation diesel ? La question pourra se concentrer sur le 1er véhicule du ménage.

Variable X : **diesel** (variable à créer selon `ENERGIE1 = 3` ou `ENERGIE1 <> 3`)

Variable Y : **TYPE_HAB**

Variable Z : **taillemng**

¹BTT: le Budget Temps de Transport est la somme des durées des déplacements réalisés durant une journée par un individu

6. Etudier les facteurs associés à l'usage fréquent (tous les jours) des transports en commun urbain (**frequ**tcu)

Variable X : **frequ**tcu (variable à créer selon **frequ**tcu = 1 ou non)

Variable Y : **permis**

Variable Z : **taillemng**

Management des Données

Les opérations sur les données sont en deux parties

1. La première partie décrit la préparation des données réalisée (*hors du projet*) pour constituer la base de données du projet. Elle présente les opérations, les scripts appliqués avant l'exercice. *Vous ne pouvez pas exécuter ces opérations.*
2. La seconde partie présente les opérations qui sont à réaliser (*dans le projet*) pour constituer les données pour le projet. *Vous devez exécuter ces opérations* et éventuellement les adapter pour votre sujet.

Partie 1 - Preprocessing

Chargement de la base de données et exploration

```
library(tidyverse)
```

Chargement de la base

Cette section les opérations qui ont permis la mise à disposition de la version des données pour le TP.

La base est une extraction de la moitié des ménages interrogés dans l'enquête.

```
# Chargement de la base d'origine
# load("allgre_TR.RData")
# # Extraction de la moitié des ménages pour travail étudiants
# set.seed(123)
# allgre$PB <- round( ave( sample( x= c(0,1), prob= c(0.5, 0.5), size =
#       length(allgre$id_men),replace=T) , allgre$id_men, FUN= mean) )
# table(allgre$PB)
# allgre.PB <- allgre[allgre$PB ==1,]
# # Sauvegarde de la nouvelle base
# # Exclusion des variables sans intérêt
# allgre.PB_V2 <- allgre.PB[, -c( 5 ,10 ,47 ,49 ,51 ,64 ,81 ,97 ,98 ,101 ,124:136, 139:148, 150, 151)]
# save(allgre.PB_V2, file = "allgre.PB_V2.RData")
```

Partie 2 - Data processing

Vous devriez commencer ici

Vous pouvez télécharger les données du tp et les charger dans R avec :

```
# Chargement de la nouvelle base PB
load("allgre.PB_V2.RData")
table(allgre.PB_V2$nbnd)
```

```
##
##      0      1      2      3      4      5      6      7      8      9     10     11     12     13     14     15
## 1277    91 4023 1967 7037 3649 4063 2386 2063 1413   901   618   412   251   197    75
##      16      17      18      22
##     146      53      36      44
```

```
# Verif des NO_DEPL
table(allgre.PB_V2$NO_DEPL)
```

```
##
##      0      1      2      3      4      5      6      7      8      9     10     11     12     13     14     15
```

```
## 1277 7025 6795 4745 4152 2392 1665 990 648 394 237 144 91 55 36 21
## 16 17 18 19 20 21 22
## 16 7 4 2 2 2 2
```

```
# Vérif_id_depl
table(is.na(allgre.PB_V2$id_depl))
```

```
##
## FALSE TRUE
## 29425 1277
```

```
# Nouveau id_depl
allgre.PB_V2$id_depl <- allgre.PB_V2$id_pers * 10 + allgre.PB_V2$NO_DEPL
table(is.na(allgre.PB_V2$id_depl))
```

```
##
## FALSE
## 30702
```

```
table(allgre.PB_V2$nbd)
```

```
##
## 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
## 1277 91 4023 1967 7037 3649 4063 2386 2063 1413 901 618 412 251 197 75
## 16 17 18 22
## 146 53 36 44
```

Description du contenu de la base de données

```
Ncol <- 7;
v <- Ncol * (Nlin <- ceiling(length(names(allgre.PB_V2))/Ncol )) - length(names(allgre.PB_V2))
mat <- matrix(c(names(allgre.PB_V2), rep("",v)), ncol = Ncol,
              dimnames = list(1:Nlin,c("col1", "col2", "col3", "col4", "col5", "col6", "col7"))
            )
# contenu base merge_4
knitr::kable(mat, digits = 2, caption = "Table: Nom des variables de la base *allgre.PB*")
```

Table 1: Table: Nom des variables de la base *allgre.PB*

col1	col2	col3	col4	col5	col6	col7
tir	PUIS_VP1	TYPE_STAT3	permis	fregtcu	minarr	durstat
NO_MEN	POSSES1	GENRE4	etabscol	fregtram	duree	autoroute
NO_PERS	LIEU_STAT1	ENERGIE4	OCCU1	fregurb	nbmodemec	abonpeage
NO_DEPL	TYPE_STAT1	AN_VP4	OCCU2	fregtransisere	prisecharge	id_men
zoneres.x	GENRE2	PUIS_VP4	csp	fregter	D12	id_pers
jourdepl	ENERGIE2	POSSES4	ABO_TC	situveil	D13	id_depl
TYPE_HAB	AN_VP2	LIEU_STAT4	VAL_ABO	zoneres.x.1	zoneres.y.1	id_traj
TYPE_OCU	PUIS_VP2	TYPE_STAT4	travdom	motifor	NO_TRAJ	nb_pers
Gare2	POSSES2	NB_velo	zonetrav	motoracc	TPS_MAP_DEP	nbd
Gare5	LIEU_STAT2	NB_2Rm	dispopv	zoneorig	mode	ntraj
telefon	TYPE_STAT2	COEF_MNG	PBM_STAT	heuredep	ZONE_D_TRAJ	btt
annuaire	GENRE3	zoneres.y	STAT_TRAV	mindep	ZONE_A_TRAJ	mode_V2
internet	ENERGIE3	sexe	fqvelo	motifdes	TPS_MAP_ARV	Couteff
VP_DISPO	AN_VP3	lien	FQ2R1	motdeacc	NUM_VEH	mode_depl_ag
GENRE1	PUIS_VP3	age	FQ2R2	nbarret	NB_OCCU	
ENERGIE1	POSSES3	TEL_PORT	fqvpcond	zonedest	LIEU_STAT	

col1	col2	col3	col4	col5	col6	col7
AN_VP1	LIEU_STAT3	mail	fqvppass	heurearr	NAT_STAT	

Description des variables et labels

Voir le dictionnaire des variables (fichier: DICO_VAR_EMD_GRE2010.xls)

tir: Numéro de tirage de l'observation (souvent proche du n° de zone de résidence)

NO_MEN: Numéro du ménage

NO_PERS: Numéro de la personne dans le ménage

NO_DEPL: Numéro du déplacement de la personne

zoneres.x: Numéro de la zone de résidence (voir fichier de correspondance)

jourdepl: Jour du déplacement

TYPE_HAB: Type de logement du ménage

TYPE_OCU: Type d'occupation de la personne

Gare2: N° de département de la gare sncf de référence

Gare5: Code postal de la gare sncf de référence

telefon: Disposition d'un téléphone

annuaire: Présent dans l'annuaire

internet: Disposition d'une connexion internet

VP_DISPO: Nombre de voiture particulière disponible dans le ménage

GENRE1: Type de voiture pour la 1ère voiture du ménage

ENERGIE1: Type de carburant de la 1ère voiture du ménage

AN_VP1: Année de mise en circulation de la 1ère voiture du ménage

PUIS_VP1: Puissance de la 1ère voiture du ménage

POSSES1: Type de propriété de la 1ère voiture du ménage

LIEU_STAT1: Lieu de stationnement de la 1ère voiture du ménage

TYPE_STAT1: Type de stationnement de la 1ère voiture du ménage

GENRE2, ENERGIE2, AN_VP2, PUIS_VP2, POSSES2, LIEU_STAT2, TYPE_STAT2, GENRE3, ENERGIE3, AN_VP3, PUIS_VP3, POSSES3, LIEU_STAT3, TYPE_STAT3, GENRE4, ENERGIE4, AN_VP4, PUIS_VP4, POSSES4, LIEU_STAT4, TYPE_STAT4: idem pour les voitures n°2, 3 et 4 du ménage

NB_velo: Nombre de vélos du ménage

NB_2Rm: Nombre de 2 roues motorisées du ménage

COEF_MNG: coefficient de redressement associé au ménage

zoneres.y: N° de zone de résidence

sexe: Sexe

lien: Lien avec la personne de référence du ménage

age: Age

TEL_PORT: Possession d'un téléphone portable

mail: Possession d'une adresse mail

permis: Possession du permis de conduire

etabscol: Dernier établissement scolaire fréquenté

OCCU1: Occupation principale

OCCU2: Autre occupation

csp: Catégorie Socio-professionnelle (PCS)

ABO_TC: Possession d'un abonnement de Transport en commun en général

VAL_ABO: Validité de l'abonnement TC hier

travdom: Travail ou étude à domicile

zonetrav: Zone du lieu de travail ou d'étude (Occupation principale)

dispovp: Disposition d'une voiture en générale (Déplacement domicile-travail ou études)

PBM_STAT: Problèmes de stationnement en général (sur lieu de travail ou d'études)

STAT_TRAV: Difficultés de stationnement sur lieu de travail ou d'études

fqvelo: Fréquence d'utilisation bicyclette
 FQ2R1: Fréquence d'utilisation 2 roues à moteurs (type 1 ?)
 FQ2R2: Fréquence d'utilisation 2 roues à moteurs (type 2 ?)
 fqvpcond: Fréquence d'utilisation Voiture en conducteur
 fqvppass: Fréquence d'utilisation Voiture en passager
 freqtcu: Fréquence d'utilisation Réseau urbain tag
 freqtram: Fréquence d'utilisation Tramway uniquement
 freqrurb: Fréquence d'utilisation Autre réseau urbain
 freqtransisere: Fréquence d'utilisation Réseau Cars Transisère
 freqter: Fréquence d'utilisation TER
 situveil: Situation de la personne la veille
 zonerex.x.1: N° Zone de résidence
 motifor: Motif à l'origine
 motoracc: Motif à l'origine de la personne accompagnée
 zoneorig: N° Zone origine du déplacement
 heuredep: Heure de départ
 mindep: Minute de départ
 motifdes: Motif à la destination
 motdeacc: Motif à la destination de la personne accompagnée
 nbarret: Nombre d'arrêts dans la tournée
 zonedest: N° Zone de destination
 heurearr: Heure d'arrivée
 minarr: Minute d'arrivée
 duree: Durée
 nbmodemec:: Nombre de modes mécanisés utilisés pour faire le déplacement
 prisecharge: Prise en charge du coût du déplacement
 D12: Distance à vol d'oiseau (en mètres)
 D13: Distance recalculée (en mètres)
 zonerex.y.1: N° zone de résidence
 NO_TRAJ: Numéro du trajet dans le déplacement
 TPS_MAP_DEP: Temps de marche à pied au départ
 mode: Mode de transport 'fin' du trajet ZONE_D_TRAJ: Zone fine de départ du mode mécanisé
 ZONE_A_TRAJ: Zone fine d'arrivée du mode mécanisé
 TPS_MAP_ARV: Temps de marche à pied à l'arrivée
 NUM_VEH: Numéro du véhicule utilisé
 NB_OCCU: Nombre d'occupant du véhicule
 LIEU_STAT: Lieu de stationnement
 NAT_STAT: Nature du stationnement
 durstat: Durée de recherche du stationnement
 autoroute: Utilisation autoroute à péage
 abonpeage: Abonnement d'autoroute
 id_men: Identifiant unique Ménage
 id_pers: Identifiant unique Personne
 id_depl: Identifiant unique Déplacement
 id_traj: Identifiant unique Trajet
 nb_pers: Nombre de personnes dans le ménage
 nbd: Nombre de déplacements dans la journée de la personne
 ntraj: Nombre de trajets dans le déplacement
 btt: Budget temps de transport (somme des durées de tous les déplacements de la personne dans sa journée)
 mode_V2: Mode de déplacement agrégé du trajet
 Couteff: Coût de déplacement estimé du trajet
 mode_depl_ag: Mode de déplacement agrégé au niveau déplacement

Structure de la base de données

Il s'agit d'une base de données dite 'hiérarchique'. Elle est initialement constituée de 4 fichiers :

1. Fichier *Ménage*
2. Fichier *Personne*
3. Fichier *Déplacement*
4. Fichier *Trajet*

Dans les enquêtes, les déplacements sont constitués comme une suite de trajets ou d'étapes effectués pour un certain motif. Pour un déplacement, il peut y avoir plusieurs trajets décrits, lorsque le déplacement est réalisé en plusieurs étapes (par ex: un trajet en vélo du domicile à la gare, un trajet en train, puis un trajet en vélo de la gare au lieu de travail). Cela est représenté dans la base de données fournies par plusieurs lignes pour un même déplacement.

Sur le même principe hiérarchique: un individu peut réaliser plusieurs déplacements dans sa journée, et un ménage peut être composé de plusieurs individus.

La base de données fournies est la plus 'fine', puisqu'elle décrit toutes les informations ménages, individus et déplacements pour chaque trajet. Il y a donc autant de lignes que de trajet.

La structure est gérée par les numéros identifiants des ménages, individus, déplacements et trajets :

- **id_men**: Identifiant unique Ménage
- **id_pers**: Identifiant unique Personne
- **id_depl**: Identifiant unique Déplacement
- **id_traj**: Identifiant unique Trajet

Le travail demandé dans ce tp peut nécessiter de travailler, par exemple sur une base d'individus au lieu des trajets. Pour cela il conviendra de réduire la base de données (réduire le nombre de lignes) pour ne conserver que des informations au niveau individuel (une ligne par individu).

Les éléments de scripts suivants permettent de réduire la base de données au niveau déplacement, personne et ménage.

Ils donnent aussi quelques indications sur la façon d'agréger l'information disponible à un niveau inférieur.

Définition de nouvelles variables

Personne Immobile

```
allgre.PB_V2$UN <- as.numeric(allgre.PB_V2$nbnd !=0)
table(as.numeric(allgre.PB_V2$UN))
```

```
##
##      0      1
## 1277 29425
```

Redéfinition des numéro de zone

```
allgre.PB_V2$DEST <- round(allgre.PB_V2$zonedest/1000)
head(allgre.PB_V2$DEST)
```

```
## [1] 101 101 103 101 102 101
```

```
allgre.PB_V2$ORIG <- round(allgre.PB_V2$zoneorig/1000)
head(allgre.PB_V2$ORIG)
```

```
## [1] 101 101 101 103 101 102
```

```
table(allgre.PB_V2$ORIG)
```

```
##
## 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116
## 813 593 486 314 289 580 172 242 257 377 181 446 433 195 380 357
## 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132
## 192 243 560 731 139 767 511 120 249 270 334 568 221 241 499 385
## 133 134 135 136 137 138 139 140 141 142 143 201 202 203 204 205
## 172 350 481 144 256 139 306 241 180 160 222 377 249 333 231 255
## 301 302 303 304 305 306 307 308 309 310 311 312 313 314 401 402
## 189 273 253 297 403 293 231 166 205 273 183 200 314 250 238 212
## 403 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515
## 312 1122 159 103 342 333 247 229 205 203 141 194 257 343 137 345
## 516 517 518 519 601 602 603 701 801 802 803 804 805 806 901 902
## 206 243 205 195 220 400 159 261 205 138 409 184 210 180 200 336
## 903 990
## 263 518
```

Distance parcourue moyenne par zone

```
mean(allgre.PB_V2$D12, na.rm = T); mean(allgre.PB_V2$D13, na.rm = T)
```

```
## [1] 4629.925
```

```
## [1] 6.73701
```

```
summary(allgre.PB_V2$D12)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##         0      715    2078    4630    5402   71356    2243
```

```
# nombre de distance 'NA' par zone
```

```
table(allgre.PB_V2$zoneres.y[is.na(allgre.PB_V2$D12)==1])
```

```
##
```

```
##      1      2      3      4      5      6      7      8      9     10     11     12     13     14     15     16     17     18     19     20
## 663 616 349 100 123  58  62  70  16  11  17  25  23   7  17   8  18   1   9   3
##  21  22  23  24  25  26  29  30  51  52
##  14   7   3   5  10   1   1   1   3   2
```

```
tapply(allgre.PB_V2$D12, as.factor(allgre.PB_V2$zoneres.y), mean, na.rm = TRUE)
```

```
##           1           2           3           4           5           6           7           8
## 4331.151 4383.808 4828.676 4884.900 4875.023 6120.861 5658.047 4439.104
##           9          10          11          12          13          14          15          16
## 8618.982 3655.540 7179.919 5220.047 7438.054 6879.554 10610.741 7656.462
##          17          18          19          20          21          22          23          24
## 4245.554 8160.481 3257.056 3302.265 5103.297 4204.639 1848.350 2472.000
##          25          26          27          29          30          51          52          59
## 4543.872 1817.467 5359.750      NaN 6926.115 5405.306 9644.750 2307.200
```

Définition de quelques variables

```
## GRENOBLE 2010      :
#####
table(allgre.PB_V2$OCCU1 )
```

```
##
##      1      2      3      4      5      6      7      8      9
## 11671 3168  238 1482 5404 1055 5402 1292  530
```

```
allgre.PB_V2$statut2 <- factor(allgre.PB_V2$OCCU1 ,
                              labels=c("TRAVAIL A TEMPS PLEIN", "TRAVAIL A TEMPS PARTIEL", "FORMATION STAGE", "E
table(allgre.PB_V2$statut2)
```

```
##
##      TRAVAIL A TEMPS PLEIN TRAVAIL A TEMPS PARTIEL      FORMATION STAGE
##              11671              3168              238
##              ETUDIANT              SCOLAIRE              CHOMEUR
##              1482              5404              1055
##              RETRAITE              RESTE AU FOYER              AUTRE
##              5402              1292              530
```

```
#####
levels(allgre.PB_V2$statut2) <- c("ACTIF", "ACTIF", "ACTIF", "ETUDIANT", "SCOLAIRE", "CHOMEUR", "RETRAITE
table(allgre.PB_V2$statut2)
```

```
##
##      ACTIF      ETUDIANT      SCOLAIRE      CHOMEUR      RETRAITE
##      15077      1482      5404      1055      5402
## RESTE AU FOYER      AUTRE
##      1292      530
```

```
#####
allgre.PB_V2$cspgroup <- trunc(allgre.PB_V2$csp / 10)
allgre.PB_V2$cspgroup[allgre.PB_V2$pcs %in% c(81,82)] <- 7
allgre.PB_V2$cspgroup[is.na(allgre.PB_V2$cspgroup)==T] <- 9
table(allgre.PB_V2$cspgroup)
```

```
##
##      1      2      3      4      5      6      8      9
## 399 1312 5161 5408 5930 3850 7936 706
```

Constitution des 3 bases

Constitution de la base *Déplacements*

```
# BASE DEPLACEMENTS
## exemple de calcul : nombre de trajets par déplacement
allgre.PB_V2$nbt2 <- ave( allgre.PB_V2$NO_TRAJ, allgre.PB_V2$id_depl, FUN = max)
addmargins( table(allgre.PB_V2$nbt2) )
```

```
##
##      0      1      2      3      4      Sum
## 6743 20332 1691   628   12 29406
```

```
## Réduction au déplacement : ATTENTION On ejecte les immobiles
allgreD <- distinct(allgre.PB_V2, id_depl, .keep_all = T)
# ou avec subset(allgre.PB_V2, allgre.PB_V2$NO_TRAJ == allgre.PB_V2$ntraj )
# nb trajet
addmargins( table(allgreD$nbt2) )
```

```
##
##      0      1      2      3      4      Sum
## 6727 20093   840   208      3 27871
```

```
# Par mode
table(allgreD$mode_depl_ag)
```

```
##
## Autre  MAP  TCIU  TCU   VP
##  2038  6792   332 1791 16937
```

```
# durée
addmargins( table(allgre.PB_V2$duree))
```

```
##
##      1      2      3      4      5      6      7      8      9     10     11     12     13
##  225     708    562    105   6479    66    208    139    29   5744    14     70     57
##  14      15     16     17     18     19     20     21     22     23     24     25     26
##  24   4464     12     21     26     5  2439      2     30     31     14    982      7
##  27     28     29     30     32     33     34     35     36     37     38     39     40
##  12     14      6  2641      2      6      2   456      2     21     22     10   613
##  41     42     43     44     45     46     47     48     50     52     53     55     56
##      2     10     11      2   860      1      9      7   252      2      5   135      2
##  57     58     59     60     62     63     65     66     67     68     69     70     73
##      8      2      1   725      5      2     76      3      5      7      3   113      2
##  74     75     77     78     80     82     83     85     90     95     97     98   100
##      2   198      2      6     66      1      3     27   209     14      1      1     26
##  105   110   112   113   115   120   122   125   127   128   130   135   140
##  54     12      2      2     14     62      2      6      1      1      4      9     13
##  145   150   155   157   158   165   170   180   190   195   200   205   210
##      2     31      2      1      1      6      1     22      1      4      2      2     18
##  220   225   230   240   245   250   255   270   280   292   293   300   325
##      8      4      4     10      3      1      7     12      2      1      1      2      2
##  330   340   345   369   405   480   630   750      Sum
##      3      2      2      3      1      1      1      4 29425
```

```
table(is.na(allgreD$duree))

##
## FALSE TRUE
## 27890 1258

table(cut(allgreD$duree, breaks=seq(0, max(allgreD$duree, na.rm = T), 30)), useNA = "always")

##
## (0,30] (30,60] (60,90] (90,120] (120,150] (150,180] (180,210] (210,240]
## 24544 2529 510 144 55 30 23 19
## (240,270] (270,300] (300,330] (330,360] (360,390] (390,420] (420,450] (450,480]
## 15 5 4 4 1 1 0 1
## (480,510] (510,540] (540,570] (570,600] (600,630] (630,660] (660,690] (690,720]
## 0 0 0 0 1 0 0 0
## (720,750] <NA>
## 4 1258

# nb déplacements
table(allgre.PB_V2$nbd)

##
## 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
## 1277 91 4023 1967 7037 3649 4063 2386 2063 1413 901 618 412 251 197 75
## 16 17 18 22
## 146 53 36 44
```

Constitution de la base *Individus*

```
#####
# BASE INDIVIDUS
## exemple de calcul : durée quotidienne des déplacements d'un individu
allgreD$btt2 <- ave( allgreD$duree, allgreD$id_pers, FUN = sum)
#addmargins( table(allgreD$btt2) )
## Reduction
allgreI <- distinct(allgreD, id_pers, .keep_all = T)
# subset(allgreD, allgreD$NO_DEPL == allgreD$nbd )
#addmargins( table(allgreI$btt2) )

summary(allgreI$btt); summary(allgreI$btt2)

## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's
## 1.00 40.00 65.00 91.13 110.00 1110.00 1258

## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's
## 1.00 38.00 65.00 80.38 105.00 870.00 1258

addmargins( table(allgreI$nbd) )

##
## 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
## 1258 72 1766 616 1672 700 662 333 253 156 89 56 33 19 14 5
## 16 17 18 22 Sum
## 9 3 2 2 7720

table(is.na(allgreI$nbd))
```

```
##
## FALSE
## 7720
```

Remarque: pour le calcul du temps quotidien de déplacement (BTT) les éléments *btt* et *btt2* diffèrent en raison de pré-traitement réalisés lors du calcul de *btt* (correction d'erreurs sur les horaires déclarés essentiellement)

Constitution de la base *Ménages*

```
#####
# BASE MENAGE
## exemple de calcul : taille du ménage et permis-ménage
allgreI$taillemng <- ave(allgreI$NO_PERS, allgreI$id_men, FUN = max)
allgreI$Permis_mng <- ave(allgreI$permis, allgreI$id_men, FUN = max)
addmargins( table(allgreI$taillemng) )
```

```
##
##      1      2      3      4      5      6      7      8 Sum
## 887 2388 1295 1931 959 190 54 16 7720
```

```
## Reduction
allgreM <- distinct(allgreI, id_men, .keep_all = T)
#subset(allgreI, allgreI$NO_PERS == allgreI$taillemng )
addmargins( table(allgreM$taillemng) )
```

```
##
##      1      2      3      4      5      6      7      8 Sum
## 887 1194 433 487 193 32 8 2 3236
```

```
addmargins( table(allgreM$Permis_mng) )
```

```
##
##      1      2      3 Sum
## 1730 1026 111 2867
```

```
table(allgreI$freqtcu)
```

```
##
##      1      2      3      4      5
## 625 545 620 1530 3869
```

```
table(allgreI$nbnd)
```

```
##
##      0      1      2      3      4      5      6      7      8      9      10      11      12      13      14      15
## 1258 72 1766 616 1672 700 662 333 253 156 89 56 33 19 14 5
## 16 17 18 22
## 9 3 2 2
```

```
table(allgreI$freqter)
```

```
##
##      1      2      3      4
## 6462 706 98 3
```

Partie 3 - Sauvegarde des Data

Vous pouvez maintenant sauvegarder les nouveaux tableaux de données qui ont été créés.

*Pour cela, vous devez modifier l'option du chunk eval= à TRUE

```
allgreTP <- allgre.PB_V2
save(allgreTP, file = "allgreTP.RData")
allgreD <- allgreD[, c( c(1:4), c(75:105), 106:109, 112, 114:116, 117:119)]
allgreI <- allgreI[, c( c(1:3), c(46:74) , 106, 107, 111, 113, 117, 120, 121 )]
allgreM <- allgreM[,c( c(1,2), c(5:45), 106:109, 110, 124,125 )]
save(allgreD, file = "allgreD.RData")
save(allgreI, file = "allgreI.RData")
save(allgreM, file = "allgreM.RData")
```

Partie - Représentation SIG

Pour réaliser cette partie sur vos postes de travail, il faut modifier les options de chunks : eval=TRUE

```
# chargement des packages nécessaires
library(igraph)

## Warning: package 'igraph' was built under R version 3.6.3
##
## Attaching package: 'igraph'
##
## The following objects are masked from 'package:dplyr':
##
##   as_data_frame, groups, union
##
## The following objects are masked from 'package:purrr':
##
##   compose, simplify
##
## The following object is masked from 'package:tidyr':
##
##   crossing
##
## The following object is masked from 'package:tibble':
##
##   as_data_frame
##
## The following objects are masked from 'package:stats':
##
##   decompose, spectrum
##
## The following object is masked from 'package:base':
##
##   union
library(dismo)

## Loading required package: raster
## Warning: package 'raster' was built under R version 3.6.3
## Loading required package: sp
## Warning: package 'sp' was built under R version 3.6.3
##
## Attaching package: 'raster'
```

```

## The following object is masked from 'package:dplyr':
##
##   select
## The following object is masked from 'package:tidyr':
##
##   extract
library(maptools)

## Warning: package 'maptools' was built under R version 3.6.3
## Checking rgeos availability: TRUE
library(reshape2)      # transformation format long, format large

## Warning: package 'reshape2' was built under R version 3.6.3
##
## Attaching package: 'reshape2'
## The following object is masked from 'package:tidyr':
##
##   smiths
library(sp)             # objets spatiaux
library(rgdal)          # fonctions de la bibliothèque GDAL

## Warning: package 'rgdal' was built under R version 3.6.3
## rgdal: version: 1.5-12, (SVN revision 1018)
## Geospatial Data Abstraction Library extensions to R successfully loaded
## Loaded GDAL runtime: GDAL 3.0.4, released 2020/01/28
## Path to GDAL shared files: C:/Users/joly.GAEL/Documents/R/win-library/3.6/rgdal/gdal
## GDAL binary built with GEOS: TRUE
## Loaded PROJ runtime: Rel. 6.3.1, February 10th, 2020, [PJ_VERSION: 631]
## Path to PROJ shared files: C:/Users/joly.GAEL/Documents/R/win-library/3.6/rgdal/proj
## Linking to sp version:1.4-2
## To mute warnings of possible GDAL/OSR exportToProj4() degradation,
## use options("rgdal_show_exportToProj4_warnings"="none") before loading rgdal.
library(ggplot2)        # fonctions graphiques
#library(ggthemes)      # thèmes pour ggplot
library(grid)           # fonction arrow
library(cartography)    # cartographie thématique

## Warning: package 'cartography' was built under R version 3.6.3
library(RColorBrewer)    # palettes de couleurs de C. Brewer
library(dplyr)

Sectir2010<-readOGR("FOND_MAP",layer="Sectir2010")
plot(Sectir2010)
# Affichage system de coordonnees
pathToShp <- "FOND_MAP"
Sectir2010@proj4string
ogrInfo(dsn = pathToShp,layer="Sectir2010")

```


Import des fichiers shp avec coordonnées compatibles

```
newESRI <- readOGR(dsn = "G:\\MyDATA\\Test_Reunion_Distance\\FOND_MAP", layer = "NewESRI")

## OGR data source with driver: ESRI Shapefile
## Source: "G:\\MyDATA\\Test_Reunion_Distance\\FOND_MAP", layer: "NewESRI"
## with 97 features
## It has 5 fields
## Integer64 fields read as strings:  cod_sectir

plot(newESRI)
```



Contrôle des shapefiles

Les shapefiles rencontrent des erreurs (des trous *holes*) qui sont corrigés avec les outils de `cleangeo()`

Ce nettoyage est nécessaire *avant* la jointure, qui va créer de nombreux polygones

```
#####
# https://gis.stackexchange.com/questions/113964/fixing-orphaned-holes-in-r
require(maptools)
# mosp <- readShapePoly("C:\\Users\\ijoly.INRA\\Desktop\\TEst_Reunion_Distance\\FOND_MAP\\Sectir2010.shp")
plot(newESRI, border= "lightgray")
```



```
require(devtools)
```

```
## Warning: package 'devtools' was built under R version 3.6.3
```

```
## Warning: package 'usethis' was built under R version 3.6.3
```

```
## to detect potential geometry issues  
# devtools::install_github("eblondel/cleangeo")  
require(cleangeo)
```

```
## Warning: package 'cleangeo' was built under R version 3.6.3
```

```
## Warning: package 'rgeos' was built under R version 3.6.3
```

```
# exploration of sh  
report <- clgeo_CollectionReport(newESRI)  
clgeo_SummaryReport(report)  
# cleaning  
mysp.clean <- clgeo_Clean(newESRI)
```

```
## Warning in proj4string(sp): CRS object has comment, which is lost in output
```

```
# check  
report.clean <- clgeo_CollectionReport(mysp.clean)  
clgeo_SummaryReport(report.clean)  
# geometry validity  
require(rgeos)  
sapply(slot(mysp.clean, "polygons"), function(x){  
  gIsValid(SpatialPolygons(Srl = list(x)))  
})
```

```
})
```

Carte et calcul des indicateurs par zone: CARTE NBD BTT

```
DF <- allgreI[, c("nbd", "btt", "tir", "zoneres.y", "id_pers")]
# Calcul base individus des indicateurs individuels
DF_Indiv <- na.omit(DF)
addmargins(table(allgreI$zoneres.y))
DF_Indiv$BTT_moy <- ave(DF_Indiv$btt,DF_Indiv$tir, FUN = mean)
DF_Indiv$BTT_med <- ave(na.omit(DF_Indiv$btt),DF_Indiv$tir, FUN = median)
DF_Indiv$nbd_med <- ave(DF_Indiv$nbd,DF_Indiv$tir, FUN = median)
DF_Indiv$nbd_moy <- ave(DF_Indiv$nbd,DF_Indiv$tir, FUN = mean)
DF_Indiv$Maxind <- ave(DF_Indiv$id_pers,DF_Indiv$tir, FUN = max)
# Réduction à la base Zone (avec le max(ind) par TIR)
DF_Zone <- subset(DF_Indiv, DF_Indiv$id_pers == DF_Indiv$Maxind, select =c(zoneres.y,BTT_med, BTT_moy, nbd_med, nbd_moy, tir))
```

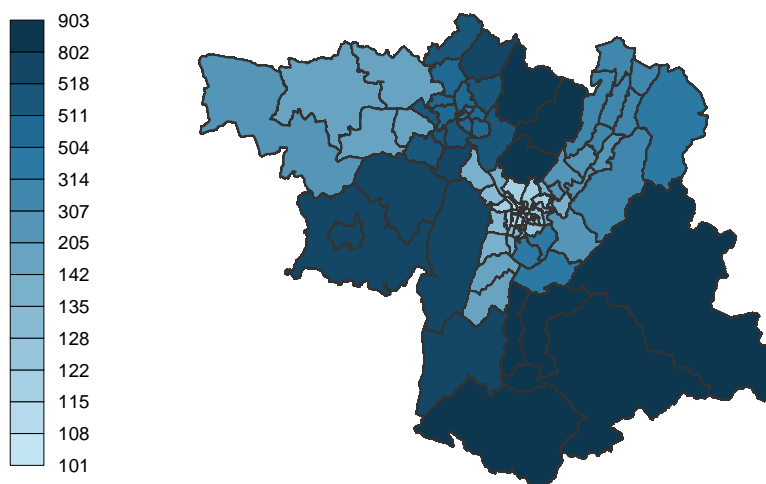
Jointure newESRI et data

```
library(sp) # the trick is that this package must be loaded!
# JOINTURE avec option duplicateGEOM=T pour avoir autant de polygones que d'observations dans data
DF_Indiv$cod_sectir <- DF_Indiv$tir
newESRI_DF <- merge(mysp.clean, DF_Indiv, by = "cod_sectir", duplicateGeoms = TRUE)
PolyG <- SpatialPolygons(mysp.clean@polygons,proj4string=mysp.clean@proj4string)
```

Carte colorée

```
choroLayer(spdf = newESRI_DF, df = newESRI_DF@data, var = "tir")
title("Mapping of the area number - not really useful")
```

Mapping of the area number – not really useful



Carte avec labels

Le package lwgeom peut être nécessaire

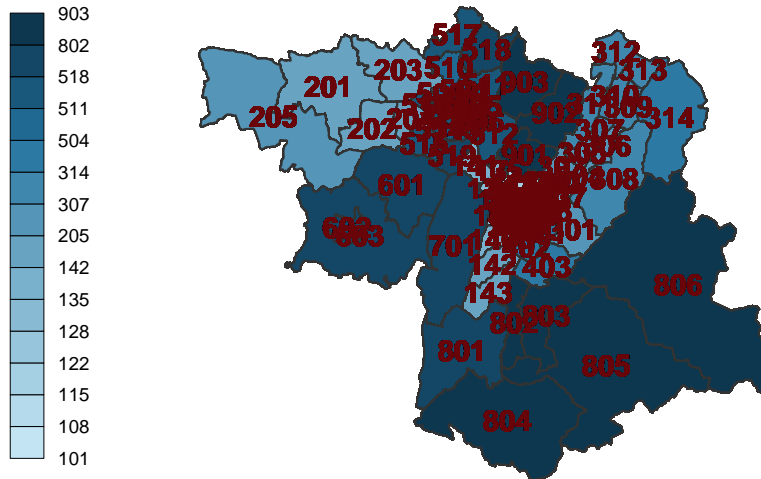
```
choroLayer(spdf = newESRI_DF, df = newESRI_DF@data, var = "tir")
# Label plot of the Mean
labelLayer(spdf = newESRI_DF, df = newESRI_DF@data, txt = "tir", col = "#690409", cex = 0.9, font = 2, sfont = "serif")

## Warning in st_centroid.sfc(x = sf::st_geometry(x), of_largest_polygon =
## max(sf::st_is(sf::st_as_sf(x), : st_centroid does not give correct centroids for
## longitude/latitude data

## Warning: package 'sf' was built under R version 3.6.3

## Linking to GEOS 3.8.0, GDAL 3.0.4, PROJ 6.3.1

## Warning in text.default(x = cc[, 1], y = cc[, 2], labels = words, cex = cex, :
## "add" n'est pas un paramètre graphique
```



Carte complète des durées quotidiennes de déplacement moyennes par zone

Ce script peut nécessiter une execution hors MARKDOWN

```
Nclasse <- 5
##### Carte colorée
choroLayer(spdf = newESRI_DF,
  df = DF_Zone,
  spdfid = "cod_sectir",
  dfid = "tir",
  var = "BTT_moy",
  nclass = Nclasse,
  col = rev(brewer.pal(n = Nclasse, "RdYlBu")),
  legend.pos = "right", legend.frame = TRUE,
  legend.title.txt = "BTT moyens par zone")
##### Ajout de label : nbd par zone
# Label creation
DF_Zone$lab <- paste(round(DF_Zone$nbd_moy,1), sep = "")
# Label plot of the Mean
#labelLayer(spdf = newESRI_DF, df = DF_Zone, txt = "lab", col = "#690409", cex = 0.9, font = 2, add=T)
##### Ajout d'éléments de présentations : Layout plot
layoutLayer(title = "BTT moyens par zone",
  sources = "EMD Grenoble 2010",
  author = "I. Joly",
  scale = 0,
  north = T,
```

```
frame = TRUE,  
col = "black",  
coltitle = "white")
```

