SABD Project

Urban Mobility Analysis

B. Agard & I. Joly

octobre 2020

Table of Content

SABD Project Description

Statement

- This project is a **3 or 4** students group work.
- All groups work on the same database, with same expectations, but different angles and problematic.
- Precisely, each group has a specific **scientific question** (from the topics list at the end of this paper - topic assignment to group is randomly determined)
- Each topic points:
  1. A variable of interest Y
  2. A Set explicit variables to be **systematically included** in the analysis X and Z
  3. A topic's specific focus to be integrated in the analysis - variables describing this focus will have to be *identified / chosen* and **justified** by the students (new variable W)
- Each topic will be analyzed based on the following approach:
  1. **Present and describe** the dataset you are using (without giving priority to one variable over another): remarkable characteristics, properties of the different variables, underlying structures, creation of variables, factorization, etc.
  2. **Summary statistics**: Produce and draw conclusions from **Univariate and bivariate** study of each of the variables:
     - *variable to be explained*
     - *imposed variables* and their contribution to the explanation of the variable of interest
     - *relevant variable to be chosen for topic's specific focus*)
  3. **Assumptions and Strategy Proposition**: Based on your *knowledge* of the topic, the **literature** (at least one relevant and well chosen paper is expected), and the former **summary statistics**:
     - *Define* and **justify** your strategy of analysis of the topic: what are the expected factors of the variable of interest?; what are your methodological choices (types of analysis, method, model,. . .)? what is your strategy (subsetting, algorithm, data handling,. . .)?
  4. The **specific focus** influence on the variable of interest will be examined. For that you will construct a new variable W that will synthesize a subset of other variables. You are asked to :
     - Select all pertinent variables related to the specific focus
     - Prepare all considered variables in order to create groups of users with similar specific focus situation
     - Cluster your dataset using those variables, clusters are stored in new variable W
     - Analyze your clusters relatively to the specific focus variables
     - Eventually iteratively improve your results reconsidering previous steps
  5. **Prediction** of variable of interest Y considering variables X, Z and W.
     - Select (with justifications) different prediction models
     - Make predictions
     - Evaluate the performance of each model.
     - Compare the different approaches

6. **Explaination** the relationships between the variable of interest and the explanatory variables you have identified
   - *Discuss* the statistical significance (or equivalent notion) of the link between variables)
   - *Interpret* the link between variables in terms of mobility behaviors and analysis
   - *Conclude* on the assumptions you have drawn in the **point 3.**

You will produce a **10 pages maximum report** (eventualy with few appendices) dealing with the following points (you will discuss them relatively to your topic):

For each point, specify if applicable if you worked on a subset of observations and / or if you deleted or created variables.

In addition, do not hesitate to collaborate with your comrades working on other variables, the issues being linked and the results of some can also be exploited by others.

The body of the report should be a synthesis of your analyzes, where you present the elements that you think are the most important. Concrete elements (values, tables, figures) should support your assertions, but do not give all the details of your analyzes in the body of the report (the appendices are made for this). For example, if you want to study the correlations of variables, indicate only the most significant in the report, and refer to the appendices for the complete matrix of correlations.

The report must be produced in Rmarkdown (or KnitR) format with all the elements allowing to **reproduce** your analysis, and returned by email (iragael.joly@grenoble-inp.fr or ijoly.research@gmail.com) on **31/01/2020** at 11:59 p.m. last deadline.

It is **imperative to ensure the reproducibility** of your work, both in the handling of the data and in the production of the report. So be sure to tell the user which packages to install, which intermediate files to run, etc.

Database description - mobility survey

Files list:

- Databases `allgreD.RData`, `allgreI.RData`, `allgreM.RData`
- Variables Dictionary **DICO_VAR_EMD_GRE2010.xls**
- File **EnonceTP_Eco3_2019.rmd** of this statement. This `Rmd` file creates files to be used in the project. It is advised to execute this file to initiate data handling and operations.
- GIS files: They are grouped in the folder: `FOND_MAP`
- GIS code are described: `Decoupage_simplifie.xlsx`

Topics and Data

Data are an extraction from mobility survey in Grenoble in 2010. Those data are available in R format: `allgreD.RData`, `allgreI.RData`, `allgreM.RData`

The data describe the trips made by members of Grenoble households in 2010 (Cerema (2013)). They provide data relating to individual and household characteristics (household structure, motorization, sex, age, professional status, etc.) and mobility practices (numbers of trips, travel times, locations of trips, modes used, reasons for trips, etc.).

The structure of the database is described in the following sections.

List of proposed topics

1. Daily travel times

Analyze the **daily travel times** (variable BTT).[1] Tips: The literature has used the DTT to define *extreme commuters*: people traveling more than 2 hours per day.

- Variable of interest Y: **btt** - Travel Time Budget
- Variable X: **jourdepl** - day of the trip
- Variable Z: **permis** - driving licence holder
- Specific focus W: influence of **housing** on your decision variable Y

2. Mode of transport choice

Analyze the 'car' choice of the mode of travel of individuals and identify the determinants of this choice among the individual or household characteristics and modes of transport.

- Variable of interest Y: **Voiture** (variable to be created according to `mode_depl_ag = 'VP'` or not)
- Variable X: **permis** - driving licence holder
- Variable Z: **age** - age
- Specific focus W: influence of **family structure** on your decision variable Y

3. Mobility behavior

Propose a characterization of the population of mobile and immobile people and identify the determinants of mobility

- Variable of interest Y: **immobil** (variable to be created according to `nbd = 0` or `nbd > 0` - `nbd` is the number of trips)
- Variable X: **age** - age
- Variable Z: **VP_DISPO** - number of available cars in the household
- Specific focus W: influence of **car ownership** on your decision variable Y

4. Household motorization?

What are the determinants of household motorization?

- Variable of interest Y: **motorisation** (variable to be created according to `VP_DISPO = 0` or `VP_DISPO> 0` - `VP_DISPO` is the number of available cars in household)
- Variable X: **taillemng** - household size
- Variable Z: **Permis_mng**: at least 1 driving licence in the household
- Specific focus W: influence of **sociodemographic household** on your decision variable Y

5. Engine Types

Study the Grenoble car fleet. What are the determinants of choosing a diesel engine? The question can focus on the 1st vehicle of the household.

- Variable of interest Y: **diesel** (variable to be created according to `ENERGIE1 = 3` or `ENERGIE1 <> 3` - `ENERGIE1`is the type of engine of the 1rst car of the household)
- Variable X: **TYPE_HAB** - type of housing
- Variable Z: **taillemng** - household size
- Specific focus W: influence of **occupation** on your decision variable Y

[1] French for TTB: Travel Time Budget is the sum of duration of all trips performed in a day by the individual

6. Frequence of use of urban public transport

Study the factors associated with frequent (daily) use of urban public transport (`freqtcu`)

- Variable of interest Y: **freqtcu** (variable to be created according to `freqtcu = 1` or not)
- Variable X: **permis** - driving licence holder
- Variable Z: **taillemng** - household size
- Specific focus W: influence of **car ownership** on your decision variable Y

7. Trips during rush hours

Study the factors associated with time of a trip during the rush hours (daily) use of urban public transport (`heuredep` and `mindep`)

- Variable of interest Y: **`heuredep` and `mindep`**
- Variable X: **motifor** - trip purpose at origin
- Variable Z: **D12** - traveled distance (distance as the crow flies)
- Specific focus W: influence of **household mobility equipment** on your decision variable Y

8. Traveled distance

Study the factors associated with traveled distance

- Variable of interest Y: `D13`
- Variable X: **motifor** - trip purpose at origin
- Variable Z: **TYPE_HAB** - type of housing
- Specific focus W: influence of **mobility accessibility and speed** on your decision variable Y

Data Operations

Data operations are in two parts.

1. First part describes data preprocessing (off the project) that were performed to create the dataframe. It shows you the operation, the scripts applied before the exercise. *You can not run these instructions*
2. Second part describes processing (on the projet) that are to be performed to manage the data for the project. *You have to run these operations* and eventualy modify them.

Part One - Preprocessing

Database loading and exploration

```
library(tidyverse)
```

Data loading This section describes former data manipulation that lead to the actual database.

The data are extract of half of the real household in the survey

```
#  echo=TRUE, eval=FALSE
# Loading of the original database
load("allgre_TR.RData")
# Extracting half of the household for confidentiality issue
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,]
# Saving
# Exclusion of non interesting variables
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")
```

Here we load the actual database

```
#  echo=TRUE, eval=FALSE
# Loading of the resulting DF
load("allgre.PB_V2.RData")
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
```

```
# Check of NO_DEPL - number of trips
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
```

```
# Check of id_depl - id of trips
table(is.na(allgre.PB_V2$id_depl))
```

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

```
# Creation of new 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 of the database content

```
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"))
              )
# content
knitr::kable( mat, digits = 2, caption = "Table: Variable names in *allgre.PB*")
```

Table 1: Table: Variable names in *allgre.PB*

col1	col2	col3	col4	col5	col6	col7
tir	PUIS_VP1	TYPE_STAT3	permis	freqtcu	minarr	durstat
NO_MEN	POSSES1	GENRE4	etabscol	freqtram	duree	autoroute
NO_PERS	LIEU_STAT1	ENERGIE4	OCCU1	freqrurb	nbmodemec	abonpeage
NO_DEPL	TYPE_STAT1	AN_VP4	OCCU2	freqtransisere	prisecharge	id_men
zoneres.x	GENRE2	PUIS_VP4	csp	freqter	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	dispovp	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	
AN_VP1	LIEU_STAT3	mail	fqvppass	heurearr	NAT_STAT	

Variables Description and labels

See variables dictionary (file: **DICO_VAR_EMD_GRE2010.xls**)

tir: Drawing number of the observation (often close to the residence zone number)
NO_MEN: Household number
NO_PERS: Number of the person in the household
NO_DEPL: Number of the person's trip
zoneres.x: Number of the area of residence (see correspondence file)
jourdepl: Day of the move
TYPE_HAB: Type of residence of the household
TYPE_OCU: Type of occupation of the person
Gare2: Department number of the reference sncf station
Gare5: Postal code of the reference sncf station
telefon: Availability of a telephone
annuaire: Present in the telephon directory
internet: Availability of an internet connection
VP_DISPO: Number of private car available in the household
GENRE1: Type of car for the 1st car of the household
ENERGIE1: Fuel type of the 1st car in the household
AN_VP1: Year of entry into service of the 1st househol car
PUIS_VP1: Power of the 1st car in the household
POSSES1: Type of ownership of the 1st car in the household
LIEU_STAT1: Parking place of the 1st car of the household
TYPE_STAT1: Type of parking for the 1st car of the household
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 for cars n ° 2, 3 and 4 of the household
NB_velo: Number of bikes in the household
NB_2Rm: Number of 2 motorized wheels in the household
COEF_MNG: adjustment coefficient associated with the household
zoneres.y: Residence zone number
sexe: Gender
lien: Link with the household reference person
age: Age
TEL_PORT: Possession of a cell phone
mail: Possession of an email address
permis: Possession of driving license
etabscol: Last school attended
OCCU1: Main occupation
OCCU2: Other occupation
csp: Socio-professional category (PCS)
ABO_TC: Possession of a Public Transport subscription in general
VAL_ABO: Validity of the TC subscription yesterday
travdom: Work or home study
zonetrav: Zone of the place of work or study (Main occupation)
dispovp: Arrangement of a car in general (Travel home-work or study)
PBM_STAT: Parking problems in general (at work or study)
STAT_TRAV: Parking difficulties at work or study place
fqvelo: Frequency of bicycle use
FQ2R1: Frequency of use 2 wheels with motors (type 1)
FQ2R2: Frequency of use 2 wheels with motors (type 2)
fqvpcond: Frequency of use Car with driver
fqvppass: Frequency of use Car as a passenger
freqtcu: Frequency of use Urban network tag
freqtram: Tramway use frequency only
freqrurb: Frequency of use Other urban network

freqtransisere: Frequency of use of the Cars Transisère network
freqter: TER use frequency
situveil: Situation of the person the day before
zoneres.x.1: Number of the area of residence
motifor: Purpose at the origin
motoracc: Purpose for the origin of the accompanied person
zoneorig: Number of the area origin of the displacement
heuredep: Start time (hour)
mindep: Start minute
motifdes: Purpose at destination
motdeacc: Purpose at destination of the accompanied person
nbarret: Number of stops in the trip
zonedest: Destination area number
heurearr: Arrival time (hour)
minarr: Arrival minute
duree: trip duration (declared)
nbmodemec: Number of mechanized mode of transport used in the trip prisecharge: Transport cost are covered
D12: traveled distance (distance as the crow flies)
D13: traveled distance
zoneres.y.1:
NO_TRAJ:
TPS_MAP_DEP: Walk time at origin
mode: Transport mode
ZONE_D_TRAJ: Area at origin of a stop
ZONE_A_TRAJ: Area at destination of a stop
TPS_MAP_ARV: Walk time at destination
NUM_VEH: Number of the vehicle
NB_OCCU: Number of occupant
LIEU_STAT: Parking place
NAT_STAT: Type of parking place
durstat: Parking duration
autoroute: Use of the highway
abonpeage: Public transport pass holder
id_men: Household id
id_pers: Person id
id_depl: Trips id
id_traj: Stop id
nb_pers: Number of persons
nbd: Number of trips
ntraj: Number of stops
btt: Daily travel time
mode_V2: transport mode version 2
Couteff: Transport cost estimation
mode_depl_ag: Transport mode aggregated version

Data Base Structure

This is a so-called 'hierarchical' database. It initially consists of 4 files:

1. File *Household*
2. File *Person*
3. File *Trip*
4. File *Route*

In surveys, trips are made up of a series of trips or stages made for a certain reason. For a trip, there may be several routes described, when the trip is made in several stages (for example: a bicycle trip from home to the station, a train trip, then a bicycle trip from the station instead of job). This is represented in the database provided by several lines for the same trip.

On the same hierarchical principle: an individual can make several trips during the day, and a household can be made up of several individuals.

The database provided is the 'finest,' since it describes all the household, individual and travel information for each trip. There are therefore as many lines as there are trips.

The structure is managed by the identifying numbers of households, individuals, trips and route:

- `id_men`: Unique Household identifier
- `id_pers`: Unique identifier Person
- `id_depl`: Unique identifier Displacement
- `id_traj`: Unique identifier of the route

The work requested in this project may require work, for example on an individual basis instead of trips. For this, it will be necessary to reduce the database (reduce the number of lines) to keep only information at the individual level (one line per individual).

The following script elements reduce the database to trip, person and household level.

They also give some guidance on how to aggregate the information available at a lower level.

New Variables Definition

Immobil person

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

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

Area number redefinition

```
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
```

Average travelled distances

```
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
```

Some variables definitions

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

Trip DataFrame: ***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
```

Person DataFrame *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
```

Remark: to calculate the daily travel time (BTT), elements `btt` and `btt2` differ due to preprocessing operations to calculate `btt` (correction of error in the declared time of trip)

Household DataFrame *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$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
##    9    3    2    2
```

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

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

Save of the 3 DataFrames

```
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")
```

Part Two - Data processing

You should start here You can download these data files and load them:

```
load("allgreD.RData")
load("allgreI.RData")
load("allgreM.RData")
```

GIS Part

```
# 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)
library(RgoogleMaps)
## Warning: package 'RgoogleMaps' was built under R version 3.6.3
library(PBSmapping)
## Warning: package 'PBSmapping' was built under R version 3.6.3
##
## ------------------------------------------------------------
## PBS Mapping 2.72.1 -- Copyright (C) 2003-2020 Fisheries and Oceans Canada
```

```
##
## PBS Mapping comes with ABSOLUTELY NO WARRANTY;
## for details see the file COPYING.
## This is free software, and you are welcome to redistribute
## it under certain conditions, as outlined in the above file.
##
## A complete user guide 'PBSmapping-UG.pdf' is located at
## C:/Users/joly.GAEL/Documents/R/win-library/3.6/PBSmapping/doc/PBSmapping-UG.pdf
##
## Packaged on 2019-03-14
## Pacific Biological Station, Nanaimo
##
## All available PBS packages can be found at
## https://github.com/pbs-software
##
## To see demos, type '.PBSfigs()'.
## ------------------------------------------------------------
```

```
library(rgeos)
```

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

```
## rgeos version: 0.5-3, (SVN revision 634)
##  GEOS runtime version: 3.8.0-CAPI-1.13.1
##  Linking to sp version: 1.4-2
##  Polygon checking: TRUE
```

```
library(cleangeo)
```

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

```
library(devtools)
```

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

```
## Loading required package: usethis
```

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

```
library(lwgeom)
```

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

```
## Linking to liblwgeom 3.0.0beta1 r16016, GEOS 3.8.0, PROJ 6.3.1
```

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

Import of shp with compatible coordinates

```
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
```

```
#ACADIA$data <- DF_Zone
plot(newESRI)
```

```
#plot(ACADIA)
```

The shapefile encouters errors (holes) that are correcte with the tools `cleangeo()`

This cleaning is needed *before* merging, which will create many polygons

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

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

Map and Calculus of area indicators: CARTE NBD BTT

```
DF <- allgreI[, c("nbd", "btt", "tir", "zoneres.y", "id_pers")]
# Calcul base individus des indicateurs individuels
DF_Indiv <- na.omit(DF)
```

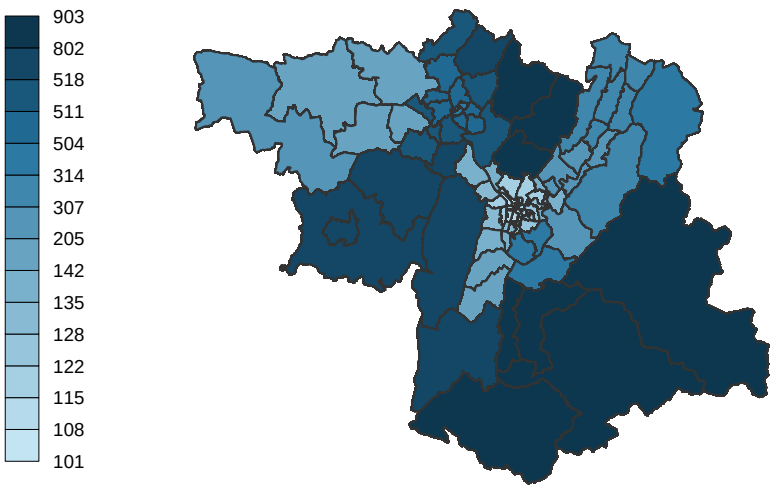
```
addmargins(table(allgreI$zoneres.y))
DF_Indiv$BTT_moy <- ave(DF_Indiv$btt,DF_Indiv$tir, FUN = mean)
DF_Indiv$BTT_med <- ave(na.omit(DF_Indiv$btt),DF_Indiv$tir, FUN = median)
DF_Indiv$nbd_med <- ave(DF_Indiv$nbd,DF_Indiv$tir, FUN = median)
DF_Indiv$nbd_moy <- ave(DF_Indiv$nbd,DF_Indiv$tir, FUN = mean)
DF_Indiv$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, n
```

Merging with newESRI and data

```
library(sp) # the trick is that this package must be loaded!
# JOINTURE avec option duplicateGEOM=T pour avoir autant de polygons 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)
```

Colored mapping

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



Map with labels

It may need to install `lwgeom`

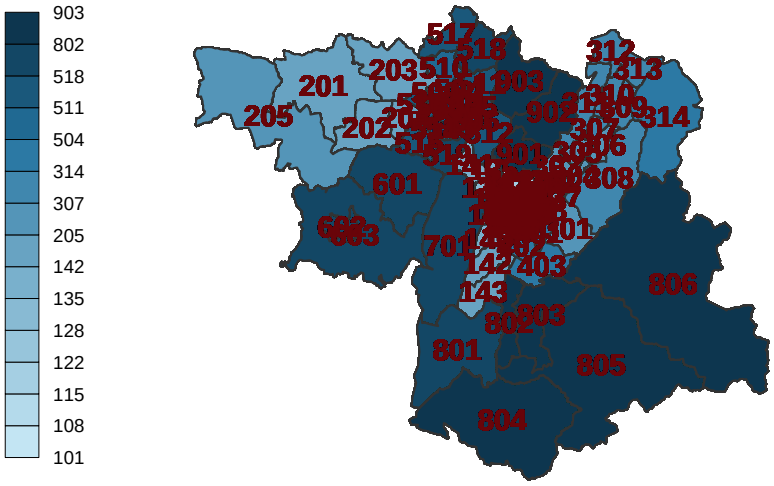
```
choroLayer(spdf = newESRI_DF, df = newESRI_DF@data, var = "tir")
# Label plot of the Mean
labelLayer(spdf = newESRI_DF, df = newESRI_DF@data, txt = "tir", col = "#690409", cex = 0.9, font = 2, a
```

```
## 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
```



References

Cerema. 2013. *Enquêtes ménages déplacements « Standard Certu »*. Cerema (ex-Certu). https://books.google.fr/books?id=nhtpmAEACAAJ.

Chang, Winston. 2013. *R Graphics Cookbook*. O'Reilly Media, Inc.

Harrell, F. E. 2013. *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis*. Springer Series in Statistics. Springer New York. https://books.google.fr/books?id=7D0mBQAAQBAJ.

Munzner, Tamara. 2015. *Visualization Analysis and Design*. CRC Press, Routledge.

Xie, Yihui. 2015. *Dynamic Documents with R and Knitr*. 2nd ed. Boca Raton, Florida: Chapman; Hall/CRC. https://yihui.name/knitr/.

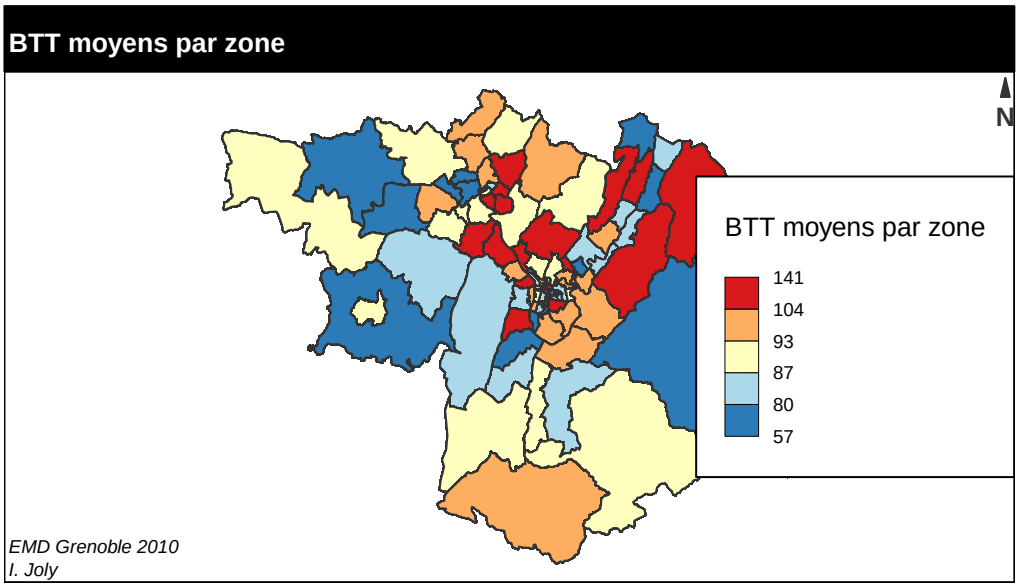
———. 2016. *Bookdown: Authoring Books and Technical Documents with R Markdown*. Boca Raton, Florida: Chapman; Hall/CRC. https://github.com/rstudio/bookdown.

Appendix

Complete map of daily travel times

The script may need to be run out of 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")
```


BTT moyens par zone
N
BTT moyens par zone
141
104
93
87
80
57
EMD Grenoble 2010
I. Joly