Effectifs_freeze

Manon Santrisse

13/10/2022

```
library(dplyr)
```

```
##
## Attachement du package : 'dplyr'
```

```
## Les objets suivants sont masqués depuis 'package:stats':
##
##     filter, lag
```

```
## Les objets suivants sont masqués depuis 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(ggplot2)
library(corrplot)
```

```
## corrplot 0.92 loaded
```

```
library(readxl)
library(FactoMineR)
library(factoextra)
```

```
## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa
```

```
library(cowplot)
library("NbClust")
library(missForest)
```

```
## Le chargement a nécessité le package : randomForest
```

```
## randomForest 4.7-1
```

```
## Type rfNews() to see new features/changes/bug fixes.
```

```
##
## Attachement du package : 'randomForest'
```

```
## L'objet suivant est masqué depuis 'package:ggplot2':
##
##     margin
```

```
## L'objet suivant est masqué depuis 'package:dplyr':
##
##     combine
```

```
## Le chargement a nécessité le package : foreach
```

```
## Le chargement a nécessité le package : itertools
```

```
## Le chargement a nécessité le package : iterators
```

```
library(writexl)
```

```
## Warning: le package 'writexl' a été compilé avec la version R 4.1.3
```

Récupération du jeu Effectifs.xlsx

```
data=read_excel('./Donnees/Effectifs_freeze.xlsx')
data
```

```
## # A tibble: 287,976 x 33
##    `Num Ctr Coll Anonyme` `Lib Entreprise An~` `Code Grp Assu~` `Condition Ven~`
##    <chr>                  <chr>                <chr>            <chr>
##  1 1-coll                 1-Lib Entreprise     NCA              NCA
##  2 1-coll                 1-Lib Entreprise     NCA              NCA
##  3 1-coll                 1-Lib Entreprise     NCA              NCA
##  4 1-coll                 1-Lib Entreprise     NCA              NCA
##  5 1-coll                 1-Lib Entreprise     NCA              NCA
##  6 1-coll                 1-Lib Entreprise     NCA              NCA
##  7 1-coll                 1-Lib Entreprise     NCA              NCA
##  8 1-coll                 1-Lib Entreprise     NCA              NCA
##  9 1-coll                 1-Lib Entreprise     NCA              NCA
## 10 1-coll                 1-Lib Entreprise     NCA              NCA
## # ... with 287,966 more rows, and 29 more variables: `Produit Anonyme` <chr>,
## #   `Date Effet Adhesion Contrat Coll` <dbl>,
## #   `Date Effet Radiation Contrat Coll` <dbl>, `Code Ape` <chr>,
## #   Departement <chr>, REGROUP_PROD_1 <chr>, REGROUP_PROD_2 <chr>,
## #   REGROUP_PROD_3 <chr>, `REGROUP_PROD_4 Anonyme` <chr>, REGROUP_PROD_5 <chr>,
## #   `Num Personne Anonyme` <chr>, `Num Ctr Indiv Anonyme` <chr>,
## #   `Date Effet Adhesion Num Personne` <dttm>, ...
```

Netoyage

Simplifications des valeurs

```
data$`Num Ctr Coll Anonyme` <- sapply(strsplit(data$`Num Ctr Coll Anonyme`,"-"),function(x){return (x[1]
data$`Lib Entreprise Anonyme`<- sapply(strsplit(data$`Lib Entreprise Anonyme`,"-"),function(x){return (
data$`Produit Anonyme`<- sapply(strsplit(data$`Produit Anonyme`,"-"),function(x){return (x[1])})
data$`Num Personne Anonyme`<- sapply(strsplit(data$`Num Personne Anonyme`,"-"),function(x){return (x[1]
data$`Num Ctr Indiv Anonyme`<- sapply(strsplit(data$`Num Ctr Indiv Anonyme`,"-"),function(x){return (x[
data$`Lien entreprise Anonyme`<- sapply(strsplit(data$`Lien entreprise Anonyme`,"-"),function(x){return
data
```

```
## # A tibble: 287,976 x 33
##    `Num Ctr Coll Anonyme` `Lib Entreprise An~` `Code Grp Assu~` `Condition Ven~`
##    <chr>                  <chr>                <chr>            <chr>
##  1 1                      1                    NCA              NCA
##  2 1                      1                    NCA              NCA
##  3 1                      1                    NCA              NCA
##  4 1                      1                    NCA              NCA
##  5 1                      1                    NCA              NCA
##  6 1                      1                    NCA              NCA
##  7 1                      1                    NCA              NCA
##  8 1                      1                    NCA              NCA
```

```
##  9 1                      1                    NCA              NCA
## 10 1                      1                    NCA              NCA
## # ... with 287,966 more rows, and 29 more variables: `Produit Anonyme` <chr>,
## #   `Date Effet Adhesion Contrat Coll` <dbl>,
## #   `Date Effet Radiation Contrat Coll` <dbl>, `Code Ape` <chr>,
## #   Departement <chr>, REGROUP_PROD_1 <chr>, REGROUP_PROD_2 <chr>,
## #   REGROUP_PROD_3 <chr>, `REGROUP_PROD_4 Anonyme` <chr>, REGROUP_PROD_5 <chr>,
## #   `Num Personne Anonyme` <chr>, `Num Ctr Indiv Anonyme` <chr>,
## #   `Date Effet Adhesion Num Personne` <dttm>, ...
```

Factorisaion de certaines variables

```
data$`Num Ctr Coll Anonyme`<-as.integer(data$`Num Ctr Coll Anonyme`)
data$`Lib Entreprise Anonyme`<-as.integer(data$`Lib Entreprise Anonyme`)
data$`Code Grp Assures`<-factor(data$`Code Grp Assures`)
data$`Condition Vente`<-factor(data$`Condition Vente`)
data$`Produit Anonyme`<-factor(data$`Produit Anonyme`)
data$`Code Ape`<-factor(data$`Code Ape`)
data$Departement<-factor(data$Departement)
data$REGROUP_PROD_1<-factor(data$REGROUP_PROD_1)
data$REGROUP_PROD_2<-factor(data$REGROUP_PROD_2)
data$REGROUP_PROD_3<-factor(data$REGROUP_PROD_3)
data$`REGROUP_PROD_4 Anonyme`<-factor(data$`REGROUP_PROD_4 Anonyme`)
data$REGROUP_PROD_5<-factor(data$REGROUP_PROD_5)
data$`Num Ctr Indiv Anonyme`<- factor(data$`Num Ctr Indiv Anonyme`)
data$`Type Assure`<-factor(data$`Type Assure`)
data$Sexe<-factor(data$Sexe)
data$`R/NR`<-factor(data$`R/NR`)
data$`Lien entreprise Anonyme`<-factor(data$`Lien entreprise Anonyme`)
data$`Numéro contrat coll Grands Comptes`<-factor(data$`Numéro contrat coll Grands Comptes`)
data$`Renégo 2020`<-as.logical(data$`Renégo 2020`)
data$`Renégo 2021`<-as.logical(data$`Renégo 2021`)
data$`Renégo 2022`<-as.logical(data$`Renégo 2022`)


data
```

```
## # A tibble: 287,976 x 33
##    `Num Ctr Coll Anonyme` `Lib Entreprise An~` `Code Grp Assu~` `Condition Ven~`
##                     <int>                <int> <fct>            <fct>
##  1                      1                    1 NCA              NCA
##  2                      1                    1 NCA              NCA
##  3                      1                    1 NCA              NCA
##  4                      1                    1 NCA              NCA
##  5                      1                    1 NCA              NCA
##  6                      1                    1 NCA              NCA
##  7                      1                    1 NCA              NCA
##  8                      1                    1 NCA              NCA
##  9                      1                    1 NCA              NCA
## 10                      1                    1 NCA              NCA
## # ... with 287,966 more rows, and 29 more variables: `Produit Anonyme` <fct>,
## #   `Date Effet Adhesion Contrat Coll` <dbl>,
## #   `Date Effet Radiation Contrat Coll` <dbl>, `Code Ape` <fct>,
## #   Departement <fct>, REGROUP_PROD_1 <fct>, REGROUP_PROD_2 <fct>,
```

```
## #   REGROUP_PROD_3 <fct>, `REGROUP_PROD_4 Anonyme` <fct>, REGROUP_PROD_5 <fct>,
## #   `Num Personne Anonyme` <chr>, `Num Ctr Indiv Anonyme` <fct>,
## #   `Date Effet Adhesion Num Personne` <dttm>, ...
```

```
#levels(data$`REGROUP_PROD_1`)
#levels(data$`REGROUP_PROD_2`)
#levels(data$`REGROUP_PROD_3`)
#levels(data$`REGROUP_PROD_4 Anonyme`)
#levels(data$`REGROUP_PROD_5`)
```

Il y a une seule catégorie pour regroup_prod 2 et 3 donc on ne va pas les étudier.

```
data=data[,-c(11,12)]
head(data)
```

```
## # A tibble: 6 x 31
##   `Num Ctr Coll Anonyme` `Lib Entreprise Ano~` `Code Grp Assu~` `Condition Ven~`
##                    <int>                 <int> <fct>            <fct>
## 1                      1                     1 NCA              NCA
## 2                      1                     1 NCA              NCA
## 3                      1                     1 NCA              NCA
## 4                      1                     1 NCA              NCA
## 5                      1                     1 NCA              NCA
## 6                      1                     1 NCA              NCA
## # ... with 27 more variables: `Produit Anonyme` <fct>,
## #   `Date Effet Adhesion Contrat Coll` <dbl>,
## #   `Date Effet Radiation Contrat Coll` <dbl>, `Code Ape` <fct>,
## #   Departement <fct>, REGROUP_PROD_1 <fct>, `REGROUP_PROD_4 Anonyme` <fct>,
## #   REGROUP_PROD_5 <fct>, `Num Personne Anonyme` <chr>,
## #   `Num Ctr Indiv Anonyme` <fct>, `Date Effet Adhesion Num Personne` <dttm>,
## #   `Date Effet Radiation Num Personne` <dbl>, `Type Assure` <fct>, ...
```

Gestion du format date

```
data=transform(data,`Date Effet Adhesion Contrat Coll` =as.character(`Date Effet Adhesion Contrat Coll`)

head(data)
```

```
##   Num.Ctr.Coll.Anonyme Lib.Entreprise.Anonyme Code.Grp.Assures Condition.Vente
## 1                    1                      1              NCA             NCA
## 2                    1                      1              NCA             NCA
## 3                    1                      1              NCA             NCA
## 4                    1                      1              NCA             NCA
## 5                    1                      1              NCA             NCA
## 6                    1                      1              NCA             NCA
##   Produit.Anonyme Date.Effet.Adhesion.Contrat.Coll
## 1               1                         20220101
## 2               1                         20220101
## 3               1                         20220101
## 4               1                         20220101
## 5               1                         20220101
## 6               1                         20220101
##   Date.Effet.Radiation.Contrat.Coll Code.Ape Departement REGROUP_PROD_1
## 1                          20501231    2932Z           @        1_ERCAC
## 2                          20501231    2932Z           @        1_ERCAC
## 3                          20501231    2932Z           @        1_ERCAC
```

```
## 4                          20501231    2932Z           @        1_ERCAC
## 5                          20501231    2932Z           @        1_ERCAC
## 6                          20501231    2932Z           @        1_ERCAC
##   REGROUP_PROD_4.Anonyme REGROUP_PROD_5 Num.Personne.Anonyme
## 1       1-Regroup_Prod_4         5_BASE                   1
## 2       1-Regroup_Prod_4         5_BASE                   2
## 3       1-Regroup_Prod_4         5_BASE                   3
## 4       1-Regroup_Prod_4         5_BASE                   4
## 5       1-Regroup_Prod_4         5_BASE                   5
## 6       1-Regroup_Prod_4         5_BASE                   6
##   Num.Ctr.Indiv.Anonyme Date.Effet.Adhesion.Num.Personne
## 1                     1                       2022-01-01
## 2                     2                       2022-01-01
## 3                     3                       2022-01-01
## 4                     4                       2022-01-01
## 5                     5                       2022-01-01
## 6                     6                       2022-01-01
##   Date.Effet.Radiation.Num.Personne Type.Assure Sexe Date.Naissance R.NR
## 1                          20501231      ENFANT    F     2005-03-02    R
## 2                          20501231      ENFANT    F     2010-05-21    R
## 3                          20501231      ENFANT    M     1999-01-08    R
## 4                          20501231      CONJOI    F     1991-08-27    R
## 5                          20501231      ENFANT    F     2017-12-02    R
## 6                          20501231      ASSPRI    M     1980-10-10    R
##   Lien.entreprise.Anonyme Numéro.contrat.coll.Grands.Comptes Indexation.2018
## 1                       1                      Grands comptes             NA
## 2                       1                      Grands comptes             NA
## 3                       1                      Grands comptes             NA
## 4                       1                      Grands comptes             NA
## 5                       1                      Grands comptes             NA
## 6                       1                      Grands comptes             NA
##   Indexation.2019 Indexation.2020 Indexation.2021 Indexation.2022
## 1              NA              NA              NA              NA
## 2              NA              NA              NA              NA
## 3              NA              NA              NA              NA
## 4              NA              NA              NA              NA
## 5              NA              NA              NA              NA
## 6              NA              NA              NA              NA
##   Indexation.2023 Renégo.2020 Renégo.2021 Renégo.2022
## 1            0.06       FALSE       FALSE       FALSE
## 2            0.06       FALSE       FALSE       FALSE
## 3            0.06       FALSE       FALSE       FALSE
## 4            0.06       FALSE       FALSE       FALSE
## 5            0.06       FALSE       FALSE       FALSE
## 6            0.06       FALSE       FALSE       FALSE
```

```
data$Date.Effet.Adhesion.Contrat.Coll<-as.Date(data$Date.Effet.Adhesion.Contrat.Coll, format = "%Y%m%d"
data$Date.Effet.Radiation.Contrat.Coll <-as.Date(data$Date.Effet.Radiation.Contrat.Coll, format = "%Y%m
data$Date.Effet.Adhesion.Num.Personne <- as.Date(data$Date.Effet.Adhesion.Num.Personne,  format = "%Y-%
data$Date.Effet.Radiation.Num.Personne <-as.Date(data$Date.Effet.Radiation.Num.Personne, format = "%Y%m
data$Date.Naissance <- as.Date(data$Date.Naissance,  format = "%Y-%m-%d")
head(data)
```

```
##   Num.Ctr.Coll.Anonyme Lib.Entreprise.Anonyme Code.Grp.Assures Condition.Vente
## 1                    1                      1              NCA             NCA
```

```
## 2                    1                      1              NCA             NCA
## 3                    1                      1              NCA             NCA
## 4                    1                      1              NCA             NCA
## 5                    1                      1              NCA             NCA
## 6                    1                      1              NCA             NCA
##   Produit.Anonyme Date.Effet.Adhesion.Contrat.Coll
## 1               1                       2022-01-01
## 2               1                       2022-01-01
## 3               1                       2022-01-01
## 4               1                       2022-01-01
## 5               1                       2022-01-01
## 6               1                       2022-01-01
##   Date.Effet.Radiation.Contrat.Coll Code.Ape Departement REGROUP_PROD_1
## 1                        2050-12-31    2932Z           @        1_ERCAC
## 2                        2050-12-31    2932Z           @        1_ERCAC
## 3                        2050-12-31    2932Z           @        1_ERCAC
## 4                        2050-12-31    2932Z           @        1_ERCAC
## 5                        2050-12-31    2932Z           @        1_ERCAC
## 6                        2050-12-31    2932Z           @        1_ERCAC
##   REGROUP_PROD_4.Anonyme REGROUP_PROD_5 Num.Personne.Anonyme
## 1       1-Regroup_Prod_4         5_BASE                    1
## 2       1-Regroup_Prod_4         5_BASE                    2
## 3       1-Regroup_Prod_4         5_BASE                    3
## 4       1-Regroup_Prod_4         5_BASE                    4
## 5       1-Regroup_Prod_4         5_BASE                    5
## 6       1-Regroup_Prod_4         5_BASE                    6
##   Num.Ctr.Indiv.Anonyme Date.Effet.Adhesion.Num.Personne
## 1                     1                       2022-01-01
## 2                     2                       2022-01-01
## 3                     3                       2022-01-01
## 4                     4                       2022-01-01
## 5                     5                       2022-01-01
## 6                     6                       2022-01-01
##   Date.Effet.Radiation.Num.Personne Type.Assure Sexe Date.Naissance R.NR
## 1                        2050-12-31      ENFANT    F     2005-03-02    R
## 2                        2050-12-31      ENFANT    F     2010-05-21    R
## 3                        2050-12-31      ENFANT    M     1999-01-08    R
## 4                        2050-12-31      CONJOI    F     1991-08-27    R
## 5                        2050-12-31      ENFANT    F     2017-12-02    R
## 6                        2050-12-31      ASSPRI    M     1980-10-10    R
##   Lien.entreprise.Anonyme Numéro.contrat.coll.Grands.Comptes Indexation.2018
## 1                       1                      Grands comptes             NA
## 2                       1                      Grands comptes             NA
## 3                       1                      Grands comptes             NA
## 4                       1                      Grands comptes             NA
## 5                       1                      Grands comptes             NA
## 6                       1                      Grands comptes             NA
##   Indexation.2019 Indexation.2020 Indexation.2021 Indexation.2022
## 1              NA              NA              NA              NA
## 2              NA              NA              NA              NA
## 3              NA              NA              NA              NA
## 4              NA              NA              NA              NA
## 5              NA              NA              NA              NA
## 6              NA              NA              NA              NA
```

```
##   Indexation.2023 Renégo.2020 Renégo.2021 Renégo.2022
## 1            0.06       FALSE       FALSE       FALSE
## 2            0.06       FALSE       FALSE       FALSE
## 3            0.06       FALSE       FALSE       FALSE
## 4            0.06       FALSE       FALSE       FALSE
## 5            0.06       FALSE       FALSE       FALSE
## 6            0.06       FALSE       FALSE       FALSE
```

Racourcis noms variables

```
Num_E=data$Num.Ctr.Coll.Anonyme
Nom_E=data$Lib.Entreprise.Anonyme
Gp_assures=data$Code.Grp.Assures
cond_vente=data$Condition.Vente
Prod=data$Produit.Anonyme
Date_adh_coll=data$Date.Effet.Adhesion.Contrat.Coll
Date_rad_coll=data$Date.Effet.Radiation.Contrat.Coll
Secteur=data$Code.Ape
REGROUP_PROD_4=data$REGROUP_PROD_4.Anonyme
Num_P=data$Num.Personne.Anonyme
Num_Fam=data$Num.Ctr.Indiv.Anonyme
Date_adh_pers=data$Date.Effet.Adhesion.Num.Personne
Date_rad_pers=data$Date.Effet.Radiation.Num.Personne
Type=data$Type.Assure
Date_Naissance=data$Date.Naissance
Lien=data$Lien.entreprise.Anonyme
Index2018=data$Indexation.2018
Index2019=data$Indexation.2019
Index2020=data$Indexation.2020
Index2021=data$Indexation.2021
Index2022=data$Indexation.2022
Index2023=data$Indexation.2023
Renego2020=data$Renégo.2020
Renego2021=data$Renégo.2021
Renego2022=data$Renégo.2022
```

Summary

Summary pour voir la tendance de répartition des variables.

```
summary(data)
```

```
##  Num.Ctr.Coll.Anonyme Lib.Entreprise.Anonyme Code.Grp.Assures Condition.Vente
##  Min.   :   1         Min.   :   1           ACT    :125100   COL    :120616
##  1st Qu.: 909         1st Qu.: 884           RET    : 22189   ACT    : 84885
##  Median :1512         Median :1458           ACT1   : 21997   RET    : 10426
##  Mean   :1709         Mean   :1645           PORT   : 18730   ACT1   :  8843
##  3rd Qu.:2437         3rd Qu.:2364           NCA    : 12554   ACC    :  8189
##  Max.   :3843         Max.   :3716           ACT4   :  9363   PORT   :  7754
##                                              (Other): 78043   (Other): 47263
##  Produit.Anonyme  Date.Effet.Adhesion.Contrat.Coll
##  641    : 10546   Min.   :1900-01-01
##  856    :  9846   1st Qu.:2013-01-01
##  972    :  8485   Median :2016-01-01
```

```
##  855    :  5722   Mean   :2014-07-22
##  1954   :  4844   3rd Qu.:2019-01-01
##  86     :  4417   Max.   :2022-10-01
##  (Other):244116
##  Date.Effet.Radiation.Contrat.Coll    Code.Ape       Departement
##  Min.   :2023-01-01                @      :102399   @      :185809
##  1st Qu.:2050-12-31                7010Z  : 16481   12     : 33457
##  Median :2050-12-31                3030Z  : 13857   66     : 14080
##  Mean   :2050-12-19                6512Z  :  8877   46     : 11989
##  3rd Qu.:2050-12-31                8810A  :  8128   11     :  9318
##  Max.   :2050-12-31                8899B  :  7852   15     :  8712
##                                    (Other):130382   (Other): 24611
##  REGROUP_PROD_1         REGROUP_PROD_4.Anonyme        REGROUP_PROD_5
##  1_ERCAC:180099   27-Regroup_Prod_4: 58043     5_BASE        :254674
##  1_ER24 : 25388   5-Regroup_Prod_4 : 36581     5_SURCOMP     : 19214
##  1_ER11 : 23335   19-Regroup_Prod_4: 19895     5_BASE_SURCOMP:  6800
##  1_CPM  : 22433   11-Regroup_Prod_4: 10666     5_OPTION_PH15 :  4689
##  1_ER66 : 12470   51-Regroup_Prod_4:  8656     5_OPTION_Sante:  1360
##  1_ER76 : 10295   39-Regroup_Prod_4:  7629     5_OPTION_IJ   :  1063
##  (Other): 13956   (Other)          :146506     (Other)       :   176
##  Num.Personne.Anonyme Num.Ctr.Indiv.Anonyme Date.Effet.Adhesion.Num.Personne
##  Length:287976        36739  :    44        Min.   :1958-10-01
##  Class :character     72573  :    42        1st Qu.:2015-12-01
##  Mode  :character     36733  :    36        Median :2018-05-01
##                       72574  :    35        Mean   :2017-03-16
##                       36977  :    32        3rd Qu.:2021-01-01
##                       72709  :    31        Max.   :2023-04-01
##                       (Other):287756
##  Date.Effet.Radiation.Num.Personne Type.Assure     Sexe
##  Min.   :2018-01-01                ASSPRI:172432   F:150628
##  1st Qu.:2021-01-01                AUTRE :    24   I:    96
##  Median :2050-12-31                CONJOI: 46898   M:137252
##  Mean   :2037-08-04                ENFANT: 68622
##  3rd Qu.:2050-12-31
##  Max.   :2050-12-31
##
##  Date.Naissance       R.NR        Lien.entreprise.Anonyme
##  Min.   :1913-02-22   NR: 15188   109    : 19609
##  1st Qu.:1959-05-02   R :272788   340    : 16133
##  Median :1977-08-04               404    : 13730
##  Mean   :1977-12-23               64     : 10627
##  3rd Qu.:1998-03-19               39     :  9213
##  Max.   :2022-09-13               69     :  8669
##                                   (Other):209995
##  Numéro.contrat.coll.Grands.Comptes Indexation.2018 Indexation.2019
##  Grands comptes: 89253              Min.   :0.00    Min.   :-0.05
##  NA's          :198723              1st Qu.:0.00    1st Qu.: 0.00
##                                     Median :0.03    Median : 0.00
##                                     Mean   :0.03    Mean   : 0.01
##                                     3rd Qu.:0.04    3rd Qu.: 0.00
##                                     Max.   :0.30    Max.   : 0.20
##                                     NA's   :94200   NA's   :67470
##  Indexation.2020  Indexation.2021 Indexation.2022 Indexation.2023
##  Min.   :-0.02    Min.   :0.00    Min.   :0.00    Min.   :0.000
```

```
##  1st Qu.: 0.00    1st Qu.:0.00    1st Qu.:0.00    1st Qu.:0.000
##  Median : 0.00    Median :0.00    Median :0.02    Median :0.085
##  Mean   : 0.01    Mean   :0.01    Mean   :0.02    Mean   :0.079
##  3rd Qu.: 0.01    3rd Qu.:0.01    3rd Qu.:0.03    3rd Qu.:0.125
##  Max.   : 0.20    Max.   :0.10    Max.   :0.20    Max.   :0.250
##  NA's   :108091   NA's   :94794   NA's   :60864   NA's   :14266
##  Renégo.2020     Renégo.2021     Renégo.2022
##  Mode :logical   Mode :logical   Mode :logical
##  FALSE:277164    FALSE:269176    FALSE:275305
##  TRUE :10812     TRUE :18800     TRUE :12671
##
##
##
##
```

Attribuation des valeurs pour chaque variable

- Gp assurés : mettre majorité
- Condition Vente : mettre majorité. Attention ! Lien entre Gp Assurés et Condition de vente ? des catégories proches, ne pas perdre le sens
- Produit : Il y a plus de 2000 produits donc on ne peut pas faire une variable par produit. Une entreprise semble avoir au maximum 3 produits et chaque produit mène à une indexation différente donc il serait intéressant de garder les 3 produits les plus représenté par ordre et leur indexation
- Date d'adhésion : pas d'intérêt / mettre année la plus vieille
- Code APE :
- Département : mettre majorité
- Regroup 1 : mettre majorité
- Regroup 4 : semble 1 entreprise = 1 regroup 4
- Regroup 5 : Créer 4 catégories (Base, Option, Asso, Surcomp) et mettre la majorité
- Num pers/ num Ind: calculer nombre de personnes et de famille. Attention ! Enlever les personnes radiées
- Date adhésion pers : pas d'intérêt
- Date radiation : voir précédement
- Type assuré : 3 catégories = variables avec leurs effectifs
- Sexe : 2 catégories = variables avec leurs effectifs
- Date de naissance : pas d'intérêt / sinon tranche d'âge
- R/RN : 2 catégories = variables avec leurs effectifs
- Num contrat grandes entreprises: faire 1 variable oui/non
- Indexation 2018/2019/2020 : faire par produit ou si 1 produit faire une moyenne ou plusieurs catégorie avec des intervalles
- Renégo 2020/2021/2022 : variable oui/non

Récupération du nombre exacte d'entreprises

```
print(max(Num_E))
```

```
## [1] 3843
```

Création d'un dataframe pour récupérer ces nouvelles informations

```
tab <- data.frame(Entreprise = integer(),
                  Lien_E = integer(),
                  Gp_ass = character(),
```

```
                Cd_vente = character(),
                Produit1= integer(),
                Porduit2= integer(),
                Produit3= integer(),
                Code_APE= character(),
                Département= character(),
                Regroup1= character(),
                Regroup4=character(),
                Regroup5=character(),
                Nb_adhérent = integer(),
                Nb_fam =integer(),
                Type_assures = character(),
                Sexe = character(),
                R_RN =character(),
                VIP = logical(),
                Ind2018 = double(),
                Ind2019 = double(),
                Ind2020 = double(),
                Renégo2020 = logical(),
                Renégo2021 = logical(),
                Renégo2022= logical())
```

Subdataset par num entreprise

Création d'une liste qui pour chaque indice regroupe les lignes correspondant à la i eme entreprise.

```
sub_entreprises <- lapply(1:3843, function(x) subset(data,Num_E==x))
```