

Effectif_freeze_Regroup

Manon Santrisse

2022-10-18

Récupération du jeu Effectifs_freeze.Rdata

```
load('../Donnees/Effectifs_freeze.Rdata')
head(effectif)
```

```
##   Num.Ctr.Coll.Anonyme Lib.Entreprise.Anonyme Code.Grp.Assures Produit.Anonyme
## 1                    1                    1          ACTIFS              1
## 2                    1                    1          ACTIFS              1
## 3                    1                    1          ACTIFS              1
## 4                    1                    1          ACTIFS              1
## 5                    1                    1          ACTIFS              1
## 6                    1                    1          ACTIFS              1
##   Date.Effet.Adhesion.Contrat.Coll Date.Effet.Radiation.Contrat.Coll Code.Ape
## 1                    2022-01-01                    2050-12-31    2932Z
## 2                    2022-01-01                    2050-12-31    2932Z
## 3                    2022-01-01                    2050-12-31    2932Z
## 4                    2022-01-01                    2050-12-31    2932Z
## 5                    2022-01-01                    2050-12-31    2932Z
## 6                    2022-01-01                    2050-12-31    2932Z
##   Departement REGROUP_PROD_1 REGROUP_PROD_4.Anonyme REGROUP_PROD_5
## 1          <NA>          ERCAC                    1          BASE
## 2          <NA>          ERCAC                    1          BASE
## 3          <NA>          ERCAC                    1          BASE
## 4          <NA>          ERCAC                    1          BASE
## 5          <NA>          ERCAC                    1          BASE
## 6          <NA>          ERCAC                    1          BASE
##   Num.Personne.Anonyme Num.Ctr.Indiv.Anonyme Date.Effet.Adhesion.Num.Personne
## 1                    1                    1          2022-01-01
## 2                    2                    2          2022-01-01
## 3                    3                    3          2022-01-01
## 4                    4                    4          2022-01-01
## 5                    5                    5          2022-01-01
## 6                    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                    TRUE                    NA
```

## 2		1		TRUE	NA
## 3		1		TRUE	NA
## 4		1		TRUE	NA
## 5		1		TRUE	NA
## 6		1		TRUE	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	NA	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	

Définition du format pour chaque variable

- Gp assurés : mettre majorité
- 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 représentés par ordre et leur indexation
- Date d'adhésion : pas d'intérêt / mettre année la plus vieille
- Code APE : mettre majorité
- 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 précédemment
- 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 catégories avec des intervalles
- Renégo 2020/2021/2022 : variable oui/non

Fonctions pour attribuer une ligne par entreprise

Dataframes de test

Garder le max représenté

Pour les variables : - Code groupe assurés - Code APE - Département - Regroup1 - Regroup4

```
max_repr <- function(data,ind) {
  frequences=as.data.frame(table(data[,ind]))
  ind_level_max=which(table(data[,ind])==max(table(data[,ind])))
  level_max=frequences[ind_level_max,1]
  return (as.character(level_max))
}
```

Test unitaire de max_repr

```
max_repr(df,1)
```

```
## [1] "b"
```

Catégories et effectifs

Pour les variables suivantes : - Type assuré - Sexe - R/RN - Regroup5

```
transpose(aggregate(E2$Type.Assure,list(E2$Type.Assure),length))
```

```
##      V1      V2
## 1 ASSPRI CONJOI
## 2      38      14
```

Date d'adhésion entreprise

Pour la variable : Date d'adhésion

```
annee_adh_E<-function(data){
  annee=year(data[,5])[1]
  return(annee)
}
```

Test unitaire

```
annee_adh_E(E1)
```

```
## [1] 2022
```

Temps d'adhésion entreprise

```
temps_adh_E<-function(data){
  if(year(data[,6])=='2050'){
    temps=as.numeric(Sys.Date()-data[,5])
  }else{
    temps=as.numeric(data[,6]-data[,5])
  }
  return(as.integer(temps))
}
#retour en jours
```

```
print(temps_adh_E(E1))
```

```
## [1] 311
```

```
print(temps_adh_E(E2))
```

```
## [1] 16931
```

Effectif personnes et familles

Pour les variables : - Num pers - Num ind

Avec retrait des personnes et familles radiées.

```
effectif_pers_fam<-function(data){  
  data_sub=subset(data, year(data[,15])=='2050')  
  nb_pers=nrow(data_sub)  
  data_sub=droplevels(data_sub)  
  nb_fam=length(levels(data_sub[,13]))  
  return (c(nb_pers,nb_fam))  
}
```

Test unitaire

```
effectif_pers_fam(E1)
```

```
## [1] 258 109
```

```
#nb pers, nb_fam
```

Temps d'adhésion moyenne des personnes

```
temps_moy_adh_P<-function(data){  
  adh=0  
  #calcul temps adhesion de toutes les personnes  
  for (i in 1:nrow(data)){  
    if(year(data[i,15])=='2050'){  
      adh=adh+as.numeric(Sys.Date()-data[i,14])  
    }else{  
      adh=adh+as.numeric(data[i,15]-data[i,14])  
    }  
  }  
  #moyenne  
  adh=adh/nrow(data)  
  return (as.integer(adh))  
}  
#en jour
```

```
temps_moy_adh_P(E2)
```

```
## [1] 4315
```

Ancienneté moyenne

```
ancien<-function(data){  
  anciennete=mean(year(data$Date.Effet.Adhesion.Num.Personne))  
  return (as.integer(anciennete))  
}
```

```
print(ancien(E1))
```

```
## [1] 2022
```

```
print(ancien(E2))
```

```
## [1] 2010
```

0/1 pour Grands comptes

```
# true or false
gd_compte<-function(data){
  return (data[1,21])
}
```

Test unitaire

```
gd_compte(E2)
```

```
## [1] FALSE
```

Récupération du nombre d'entreprises et de liens

```
effectif$Num.Ctr.Coll.Anonyme<-as.integer(effectif$Num.Ctr.Coll.Anonyme)
paste("Il y a ", max(effectif$Num.Ctr.Coll.Anonyme), "entreprises.")
```

```
## [1] "Il y a 3789 entreprises."
```

```
effectif$Num.Ctr.Coll.Anonyme<-as.factor(effectif$Num.Ctr.Coll.Anonyme)
effectif$Lien.entreprise.Anonyme<-as.integer(effectif$Lien.entreprise.Anonyme)
paste("Il y a ",max(effectif$Lien.entreprise.Anonyme)," liens")
```

```
## [1] "Il y a 1007 liens"
```

```
effectif$Lien.entreprise.Anonyme<-as.factor(effectif$Lien.entreprise.Anonyme)
```

Création d'un dataframe pour récupérer ces nouvelles informations

```
tab <- data.frame(Entreprise = integer(),
  Lien_E = integer(),
  Actifs = integer(),
  Non_actifs = integer(),
  Portabilite = integer(),
  # Produit1= integer(),
  # Produit2= integer(),
  # Produit3= integer(),
  Annee_adhesion = integer(),
  Tps_adh_entreprise = integer(),
  Max_code_ape = character(),
  Max_departement= character(),
  Max_regroup1= character(),
  Max_regroup4=integer(),
  BASE=numeric(), # en pourcentage
  OPTION = numeric(), # en pourcentage
  Nb_pers = integer(),
  Nb_fam =integer(),
  Anciennete_moy = integer(),
  Tps_adh_moy_pers = integer(),
  Asspri = integer(),# effectif ou pourcentage
  Conjoint = integer(),
```

```
Enfant = integer(),  
Femme = numeric(), # pourcentage  
Homme = numeric(), #pourcentage  
Respo = numeric(), # pourcentage  
Non_respo = numeric(), #pourcentage  
VIP = logical(),  
Ind2018 = double(),  
Ind2019 = double(),  
Ind2020 = double(),  
Ind2021 = double(),  
Ind2022 = double(),  
Ind2023 = double(),  
Renégo2020 = logical(),  
Renégo2021 = logical(),  
Renégo2022= logical()
```