

# Projet R

## Contents

Description des variables . . . . .	1
- Packages . . . . .	3
Manipulation des données . . . . .	4
Visualisation de nos variables . . . . .	8
Analyse des primes en fonction des comparateurs et des différentes formules souscrites . . . . .	12
Analyse des caractéristiques des clients ayant souscrit via les différents comparateurs . . . . .	15

Réalisé par:

Adam NGATCHOU

Enzo ROSELIA

Arnold LEKEM

## Description des variables

Fichier contrats

- IDPOL: Identifiant du contrat d'assurance
- DATEEFF: Date d'effet du contrat
- IDCLI: Identifiant du client
- IDPART: Identifiant du conjoint du client
- IDCHI1: Identifiant de l'enfant numéro 1
- IDCHI2: Identifiant de l'enfant numéro 2
- IDCHI3: Identifiant de l'enfant numéro 3
- IDCHI4: Identifiant de l'enfant numéro 4
- IDCHI5: Identifiant de l'enfant numéro 5
- IDCHI6: Identifiant de l'enfant numéro 6
- FORMU: Formule du contrat
- DATECAN: Date de résiliation
- REACANC: Raison de résiliation
- AMTPREM: Montant de prime
- RANKAGG: Rang sur le comparateur
- COMPAR: Comparateur
- MARCLI: Statut marital du client
- SEXCLI: Sexe du client
- DTNCLI: Date de naissance du client

- PROFCLI: Profession du client
- CP: Code postal du client
- SEXCJ: Sexe du conjoint
- DTNCJ: Date de naissance du conjoint
- PROFCJ: Profession du conjoint

Fichier prestations

- IDPOL: Identifiant du contrat d'assurance
- IDBENE: Identifiant du bénéficiaire (ça peut être l'assuré, le conjoint ou l'un des enfants)

Categories: catégorie de prestation

- Subcat: sous-catégorie de prestation
- Name: libellé de prestation
- Amt: montant de dépense

Durant ce projet, nous allons étudier la situation des comparateurs, et l'importance des comparateurs dans notre portefeuille d'assurance. Pour cela, nous allons analyser plusieurs caractéristiques.

```
contra=read.csv2("http://dataks.bitbucket.io/assurance/contrats.csv")
```

```
prest=read.csv2("http://dataks.bitbucket.io/assurance/prestations_bene.csv")
```

```
head(contra)
```

```
##          IDPOL    DATEEFF      IDCLI    IDPART    IDCHI1    IDCHI2 IDCHI3
## 1 2C7EAEF380 2011-02-10 7A61ADDB61      <NA>      <NA>      <NA> <NA>
## 2 F1125A72F9 2011-07-10 4E5D3AA63C 32D5615B20 2EF179B6C7 B09D69E316 <NA>
## 3 0F5BF8EA40 2011-08-10 283B50C751      <NA>      <NA>      <NA> <NA>
## 4 73FB728155 2011-08-10 2DF3A65E55      <NA>      <NA>      <NA> <NA>
## 5 21BFCE62CB 2012-01-01 9C968CCD1F      <NA>      <NA>      <NA> <NA>
## 6 0EF51BD7B0 2011-12-10 51305185AA      <NA>      <NA>      <NA> <NA>
##  IDCHI4 IDCHI5 IDCHI6 FORMU    DATECAN                                REACAN AMTPREM
## 1  <NA>  <NA>    NA  FOR1 2017-12-31 Non-payment of the premium 220.92
## 2  <NA>  <NA>    NA  FOR4 2016-12-31 Non-payment of the premium 769.44
## 3  <NA>  <NA>    NA  FOR2 2016-12-31 Non-payment of the premium 311.16
## 4  <NA>  <NA>    NA  FOR3 2016-12-31 Non-payment of the premium 583.92
## 5  <NA>  <NA>    NA  FOR4 2016-12-31 Non-payment of the premium 961.8
## 6  <NA>  <NA>    NA  FOR4 2016-12-31 Non-payment of the premium 797.76
##  RANKAGG      COMPAR  MARCLI SEXCLI    DTNCLI  PROFCLI    CP SEXCJ
## 1      1 Comparateur1 Single      M 1980-12-16 Employee 29200 <NA>
## 2      4 Comparateur1 Married    M 1968-12-22 Employee 1700    F
## 3      1 Comparateur1 Single      M 1977-06-08 Employee 45260 <NA>
## 4      2 Comparateur1 Single      M 1960-09-05 Employee 74130 <NA>
## 5      2 Comparateur1 Single      F 1951-02-19 Retired 14600 <NA>
## 6      6 Comparateur1 Single      F 1962-07-12 Employee 66600 <NA>
##          DTNCJ    PROFCJ
## 1      <NA>      <NA>
## 2 1978-11-15 Employee
## 3      <NA>      <NA>
## 4      <NA>      <NA>
## 5      <NA>      <NA>
## 6      <NA>      <NA>
```

```
head(prest)
```

```
##           IDPOL      IDBENE      Categories      Subcat
## 1 228CD89F09 D29FCC705E Ensemble des Hopitaux HOSPITALISATION
## 2 C87113477B BF4CA46421 Ensemble des Hopitaux HOSPITALISATION
## 3 29D2B59AA7 18EC43BA24 Ensemble des Hopitaux HOSPITALISATION
## 4 1EA49C7D41 5597A0EF6F Ensemble des Hopitaux HOSPITALISATION
## 5 74F33CA4BE 2C14A84AD9 Ensemble des Hopitaux HOSPITALISATION
## 6 11069421BF 5E17F9F8D6          OPTIQUE          MONTURE
##
##              Name      Amt
## 1 FRAIS D'HEBERGEMENT ET DE SOINS 486.45
## 2 FRAIS D'HEBERGEMENT ET DE SOINS 1558.00
## 3 FRAIS D'HEBERGEMENT ET DE SOINS 1297.40
## 4 FRAIS D'HEBERGEMENT ET DE SOINS 669.00
## 5 FRAIS D'HEBERGEMENT ET DE SOINS 2670.00
## 6              MONTURE OPTIQUE 50.00
```

```
#install.packages("data.table")
```

```
library("data.table")
```

## - Packages

```
## Warning: package 'data.table' was built under R version 3.6.3
```

```
library("ggplot2")
```

```
## Warning: package 'ggplot2' was built under R version 3.6.3
```

```
library(tidyverse)
```

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

```
## -- Attaching packages ----- tidyverse 1.3.0
```

```
## v tibble 3.0.4      v dplyr 1.0.2
## v tidyr 1.1.2      v stringr 1.4.0
## v readr 1.3.1      v forcats 0.5.0
## v purrr 0.3.3
```

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

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

```
## Warning: package 'readr' was built under R version 3.6.2
```

```
## Warning: package 'purrr' was built under R version 3.6.2
```

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

```
## Warning: package 'stringr' was built under R version 3.6.2
```

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

```
## -- Conflicts ----- tidyverse_conflict_
```

```
## x dplyr::between() masks data.table::between()
## x dplyr::filter() masks stats::filter()
## x dplyr::first() masks data.table::first()
## x dplyr::lag() masks stats::lag()
```

```
## x dplyr::last()      masks data.table::last()
## x purrr::transpose() masks data.table::transpose()

#install.packages("ggthemes")

#install.packages("ggsci")

library(ggthemes)

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

library("ggsci")

## Warning: package 'ggsci' was built under R version 3.6.2
  • utilisation de data.table

contra=data.table(contra)

prest=data.table(prest)
```

## Manipulation des données

```
str(contra)

## Classes 'data.table' and 'data.frame':  6993 obs. of  24 variables:
## $ IDPOL  : Factor w/ 6990 levels "000B46F609","0010C764A2",...: 1201 6569 406 3130 893 398 4562 5966
## $ DATEEFF: Factor w/ 778 levels "2011-01-12","2011-02-10",...: 2 10 13 13 20 19 18 4 16 7 ...
## $ IDCLI  : Factor w/ 6979 levels "00047E6ABB","001051A2D8",...: 3272 2054 1020 1180 4209 2134 3449 5
## $ IDPART : Factor w/ 1823 levels "0022344D31","00642A0F69",...: NA 352 NA NA NA NA 1748 1689 NA NA .
## $ IDCHI1 : Factor w/ 1980 levels "005365B5B3","007E75CB81",...: NA 346 NA NA NA NA NA NA NA NA ...
## $ IDCHI2 : Factor w/ 1101 levels "0021F9CEEB","007291596D",...: NA 779 NA NA NA NA NA NA NA NA ...
## $ IDCHI3 : Factor w/ 189 levels "011BDAC4A2","01B8C58ADE",...: NA NA NA NA NA NA NA NA NA NA ...
## $ IDCHI4 : Factor w/ 31 levels "02F9E21EC2","0AB39C9B8A",...: NA NA NA NA NA NA NA NA NA NA ...
## $ IDCHI5 : Factor w/ 2 levels "22587BB280","D2D2B0253F": NA NA NA NA NA NA NA NA NA NA ...
## $ IDCHI6 : logi  NA NA NA NA NA NA NA ...
## $ FORMU  : Factor w/ 4 levels "FOR1","FOR2",...: 1 4 2 3 4 4 4 4 1 4 ...
## $ DATECAN: Factor w/ 685 levels "2011-12-13","2011-12-18",...: 685 542 542 542 542 542 542 390 390 3
## $ REACAN : Factor w/ 3 levels "Cancel renewal policy",...: 3 3 3 3 3 3 3 3 3 3 ...
## $ AMTPREM: Factor w/ 4845 levels "1000.32","1000.44",...: 1455 4242 1977 3628 4738 4320 2990 1210 14
## $ RANKAGG: int   1 4 1 2 2 6 NA NA 1 1 ...
## $ COMPAR : Factor w/ 7 levels "Comparateur1",...: 1 1 1 1 1 1 NA NA 1 1 ...
## $ MARCLI : Factor w/ 5 levels "Cohabitant / Pacs",...: 4 2 4 4 4 4 1 2 4 4 ...
## $ SEXCLI : Factor w/ 2 levels "F","M": 2 2 2 2 1 1 2 2 2 2 ...
## $ DTNCLI : Factor w/ 5539 levels "1937-07-30","1937-11-08",...: 2750 1125 2260 420 72 553 4578 178 3
## $ PROFCLI: Factor w/ 13 levels "Company head",...: 3 3 3 3 8 3 13 12 3 12 ...
## $ CP      : int  29200 1700 45260 74130 14600 66600 39130 77000 44200 15500 ...
## $ SEXCJ   : Factor w/ 2 levels "F","M": NA 1 NA NA NA NA 1 1 NA NA ...
## $ DTNCJ   : Factor w/ 1697 levels "1940-01-16","1941-09-06",...: NA 1070 NA NA NA NA 1678 179 NA NA .
## $ PROF CJ : Factor w/ 14 levels "Company head",...: NA 3 NA NA NA NA 13 13 NA NA ...
## - attr(*, ".internal.selfref")=<externalptr>

  • La variable compareteur
  • description de la variable de la variable compareteur
```

```
str(contra[,COMPAR])
```

```
## Factor w/ 7 levels "Compareteur1",...: 1 1 1 1 1 1 NA NA 1 1 ...
```

**Remarque:** Nous remarquons qu'il y'a des "NA" dans la variable COMPAR nous devons afin de manipuler l'information extraire les informations pertinentes soit extraire les lignes de notre tables qui sont non vides.

Type caractère: OK

```
table(is.na(contra[,AMTPREM]))
```

```
##  
## FALSE  
## 6993
```

**Conclusion:** Aucun NA dans la variable AMTPREM

- Extraire les données ayant la variable COMPAR non vide

```
contra_comp=contra[COMPAR!="NA"]
```

vérification

```
table(contra_comp$COMPAR,useNA = "always")
```

```
##  
## Compareteur1 Compareteur2 Compareteur3 Compareteur4 Compareteur5 Compareteur6  
##          481          357          216          619          1726          2401  
## Compareteur7      <NA>  
##          36           0
```

- Identifions les probables problèmes de type pour le formattage des données

```
str(contra_comp)
```

```
## Classes 'data.table' and 'data.frame': 5836 obs. of 24 variables:  
## $ IDPOL : Factor w/ 6990 levels "000B46F609","0010C764A2",...: 1201 6569 406 3130 893 398 6338 6457  
## $ DATEEFF: Factor w/ 778 levels "2011-01-12","2011-02-10",...: 2 10 13 13 20 19 16 7 7 1 ...  
## $ IDCLI : Factor w/ 6979 levels "00047E6ABB","001051A2D8",...: 3272 2054 1020 1180 4209 2134 6413 5...  
## $ IDPART: Factor w/ 1823 levels "0022344D31","00642A0F69",...: NA 352 NA NA NA NA NA NA NA ...  
## $ IDCHI1: Factor w/ 1980 levels "005365B5B3","007E75CB81",...: NA 346 NA NA NA NA NA NA NA ...  
## $ IDCHI2: Factor w/ 1101 levels "0021F9CEEB","007291596D",...: NA 779 NA NA NA NA NA NA NA ...  
## $ IDCHI3: Factor w/ 189 levels "011BDAC4A2","01B8C58ADE",...: NA NA NA NA NA NA NA NA NA ...  
## $ IDCHI4: Factor w/ 31 levels "02F9E21EC2","0AB39C9B8A",...: NA NA NA NA NA NA NA NA NA ...  
## $ IDCHI5: Factor w/ 2 levels "22587BB280","D2D2B0253F": NA NA NA NA NA NA NA NA NA ...  
## $ IDCHI6: logi NA NA NA NA NA NA NA ...  
## $ FORMU : Factor w/ 4 levels "FOR1","FOR2",...: 1 4 2 3 4 4 1 4 4 1 ...  
## $ DATECAN: Factor w/ 685 levels "2011-12-13","2011-12-18",...: 685 542 542 542 542 542 390 390 390 3...  
## $ REACAN : Factor w/ 3 levels "Cancel renewal policy",...: 3 3 3 3 3 3 3 3 3 3 ...  
## $ AMTPREM: Factor w/ 4845 levels "1000.32","1000.44",...: 1455 4242 1977 3628 4738 4320 1401 1692 21...  
## $ RANKAGG: int 1 4 1 2 2 6 1 1 3 2 ...  
## $ COMPAR : Factor w/ 7 levels "Compareteur1",...: 1 1 1 1 1 1 1 1 1 1 ...  
## $ MARCLI : Factor w/ 5 levels "Cohabitant / Pacs",...: 4 2 4 4 4 4 4 4 4 4 ...  
## $ SEXCLI : Factor w/ 2 levels "F","M": 2 2 2 2 1 1 2 2 2 2 ...  
## $ DTNCLI : Factor w/ 5539 levels "1937-07-30","1937-11-08",...: 2750 1125 2260 420 72 553 3082 2163 ...  
## $ PROFCLI: Factor w/ 13 levels "Company head",...: 3 3 3 3 8 3 3 12 3 3 ...  
## $ CP : int 29200 1700 45260 74130 14600 66600 44200 15500 34500 6000 ...  
## $ SEXCJ : Factor w/ 2 levels "F","M": NA 1 NA NA NA NA NA NA NA NA ...  
## $ DTNCJ : Factor w/ 1697 levels "1940-01-16","1941-09-06",...: NA 1070 NA NA NA NA NA NA NA NA ...
```

```
## $ PROFCJ : Factor w/ 14 levels "Company head",...: NA 3 NA NA NA NA NA NA NA ...
## - attr(*, ".internal.selfref")=<externalptr>
```

**Remarque :** Le type de la variable AMTPREM qui représente les primes versées, n'est pas adéquat, nous devons changer le type en numérique.

- Variable primes

```
contra_comp$AMTPREM=as.double(contra_comp$AMTPREM)
```

- Vérification

```
class(contra_comp$AMTPREM)
```

```
## [1] "numeric"
```

Modification OK

```
table(is.na(contra_comp[,AMTPREM]),useNA = "always")
```

```
##
## FALSE <NA>
## 5836      0
```

```
table(is.na(contra_comp[,SEXCLI]),useNA = "always")
```

```
##
## FALSE <NA>
## 5836      0
```

```
table(is.na(contra_comp[,FORMU]),useNA = "always")
```

```
##
## FALSE <NA>
## 5836      0
```

```
table(is.na(contra_comp[,COMPAR]),useNA = "always")
```

```
##
## FALSE <NA>
## 5836      0
```

```
table(is.na(contra_comp[,MARCLI]),useNA = "always")
```

```
##
## FALSE <NA>
## 5836      0
```

```
summary(contra_comp)
```

##	IDPOL	DATEFF	IDCLI	IDPART
##	5DA4AB893C: 2	2018-01-01: 321	0FF7B65E3F: 2	05951BEE48: 2
##	AFB99189E1: 2	2017-01-02: 141	2CABA15A47: 2	0022344D31: 1
##	F4631B2A31: 2	2017-01-04: 122	2CE3E0F5D5: 2	00642A0F69: 1
##	000B46F609: 1	2017-01-12: 105	A0C919874D: 2	00D12ACBC6: 1
##	0010C764A2: 1	2015-01-01: 104	E1DADD36E8: 2	010E2887EE: 1
##	0014EBDBEB: 1	2014-01-01: 99	00047E6ABB: 1	(Other) :1495
##	(Other) :5827	(Other) :4944	(Other) :5825	NA's :4335
##	IDCHI1	IDCHI2	IDCHI3	IDCHI4
##	7AFDF293CA: 2	107DBA0E6D: 2	011BDAC4A2: 1	02F9E21EC2: 1
##	005365B5B3: 1	007291596D: 1	01B8C58ADE: 1	0AB39C9B8A: 1

```

## 007E75CB81: 1 00E6C55736: 1 038FC294D8: 1 0B927EE04F: 1
## 008410AEA6: 1 014233B42C: 1 0483187CA0: 1 13AE15F74A: 1
## 00877AC0FD: 1 026D232DB0: 1 05A3B4FC00: 1 21A96F7944: 1
## (Other) :1626 (Other) : 899 (Other) : 154 (Other) : 24
## NA's :4204 NA's :4931 NA's :5677 NA's :5807
## IDCHI5 IDCHI6 FORMU DATECAN
## 22587BB280: 1 Mode:logical FOR1: 368 2015-12-31: 16
## D2D2B0253F: 1 NA's:5836 FOR2: 164 2016-02-28: 11
## NA's :5834 FOR3: 118 2016-12-15: 8
## FOR4:5186 2017-01-20: 8
## 2014-01-31: 7
## (Other) :1099
## NA's :4687
## REACAN AMTPREM
## Cancel renewal policy : 234 Min. : 2
## Cancellation on renewal from the customer: 623 1st Qu.:1593
## Non-payment of the premium :1745 Median :2502
## NA's :3234 Mean :2486
## 3rd Qu.:3466
## Max. :4845
## RANKAGG COMPAR MARCLI SEXCLI
## Min. : 0.000 Comparateur1: 481 Cohabitant / Pacs : 604 F:2715
## 1st Qu.: 1.000 Comparateur2: 357 Married :1816 M:3121
## Median : 2.000 Comparateur3: 216 Separated / Divorced: 184
## Mean : 2.711 Comparateur4: 619 Single :3207
## 3rd Qu.: 3.000 Comparateur5:1726 widower : 25
## Max. :16.000 Comparateur6:2401
## NA's :3328 Comparateur7: 36
## DTNCLI PROFCLI CP
## 1973-12-19: 5 Without profession :1890 Min. : 1000
## 1977-10-31: 5 Employee :1876 1st Qu.:34970
## 1991-02-21: 5 Other :1250 Median :63380
## 1968-01-15: 4 Public administration employee: 399 Mean :58668
## 1976-08-07: 4 Retired : 129 3rd Qu.:78415
## 1981-04-19: 4 Executive : 95 Max. :95880
## (Other) :5809 (Other) : 197
## SEXCJ DTNCJ PROFCJ
## F : 888 1958-08-17: 3 Employee : 555
## M : 613 1965-11-27: 3 Without profession : 483
## NA's:4335 1970-03-03: 3 Other : 294
## 1973-03-29: 3 Public administration employee: 90
## 1973-05-02: 3 Executive : 22
## (Other) :1486 (Other) : 57
## NA's :4335 NA's :4335

```

```
str(contra_comp)
```

```

## Classes 'data.table' and 'data.frame': 5836 obs. of 24 variables:
## $ IDPOL : Factor w/ 6990 levels "000B46F609","0010C764A2",...: 1201 6569 406 3130 893 398 6338 6457
## $ DATEEFF: Factor w/ 778 levels "2011-01-12","2011-02-10",...: 2 10 13 13 20 19 16 7 7 1 ...
## $ IDCLI : Factor w/ 6979 levels "00047E6ABB","001051A2D8",...: 3272 2054 1020 1180 4209 2134 6413 5
## $ IDPART : Factor w/ 1823 levels "0022344D31","00642A0F69",...: NA 352 NA NA NA NA NA NA NA ...
## $ IDCHI1 : Factor w/ 1980 levels "005365B5B3","007E75CB81",...: NA 346 NA NA NA NA NA NA NA ...
## $ IDCHI2 : Factor w/ 1101 levels "0021F9CEEB","007291596D",...: NA 779 NA NA NA NA NA NA NA ...

```

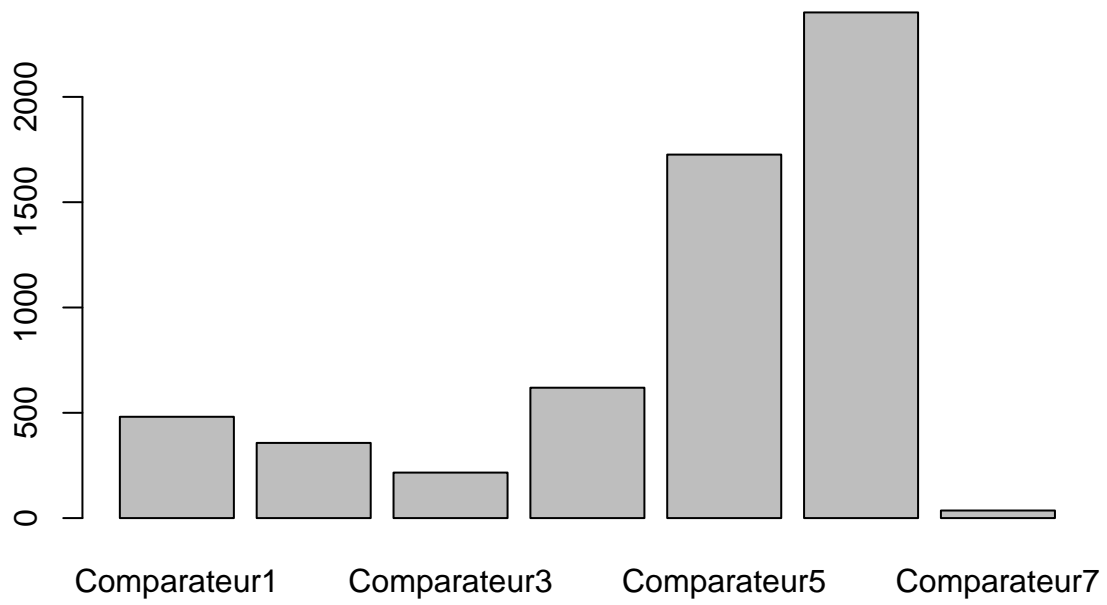
```
## $ IDCHI3 : Factor w/ 189 levels "011BDAC4A2","01B8C58ADE",...: NA NA NA NA NA NA NA NA NA NA ...
## $ IDCHI4 : Factor w/ 31 levels "02F9E21EC2","0AB39C9B8A",...: NA NA NA NA NA NA NA NA NA NA ...
## $ IDCHI5 : Factor w/ 2 levels "22587BB280","D2D2B0253F": NA NA NA NA NA NA NA NA NA NA ...
## $ IDCHI6 : logi NA NA NA NA NA NA NA ...
## $ FORMU : Factor w/ 4 levels "FOR1","FOR2",...: 1 4 2 3 4 4 1 4 4 1 ...
## $ DATECAN: Factor w/ 685 levels "2011-12-13","2011-12-18",...: 685 542 542 542 542 542 390 390 390 390 ...
## $ REACAN : Factor w/ 3 levels "Cancel renewal policy",...: 3 3 3 3 3 3 3 3 3 3 ...
## $ AMTPREM: num 1455 4242 1977 3628 4738 ...
## $ RANKAGG: int 1 4 1 2 2 6 1 1 3 2 ...
## $ COMPAR : Factor w/ 7 levels "Compareteur1",...: 1 1 1 1 1 1 1 1 1 1 ...
## $ MARCLI : Factor w/ 5 levels "Cohabitant / Pacs",...: 4 2 4 4 4 4 4 4 4 4 ...
## $ SEXCLI : Factor w/ 2 levels "F","M": 2 2 2 2 1 1 2 2 2 2 ...
## $ DTNCLI : Factor w/ 5539 levels "1937-07-30","1937-11-08",...: 2750 1125 2260 420 72 553 3082 2163 ...
## $ PROFCLI: Factor w/ 13 levels "Company head",...: 3 3 3 3 8 3 3 12 3 3 ...
## $ CP : int 29200 1700 45260 74130 14600 66600 44200 15500 34500 6000 ...
## $ SEXCJ : Factor w/ 2 levels "F","M": NA 1 NA NA NA NA NA NA NA NA ...
## $ DTNCJ : Factor w/ 1697 levels "1940-01-16","1941-09-06",...: NA 1070 NA NA NA NA NA NA NA NA ...
## $ PROF CJ : Factor w/ 14 levels "Company head",...: NA 3 NA NA NA NA NA NA NA NA ...
## - attr(*, ".internal.selfref")=<externalptr>
```



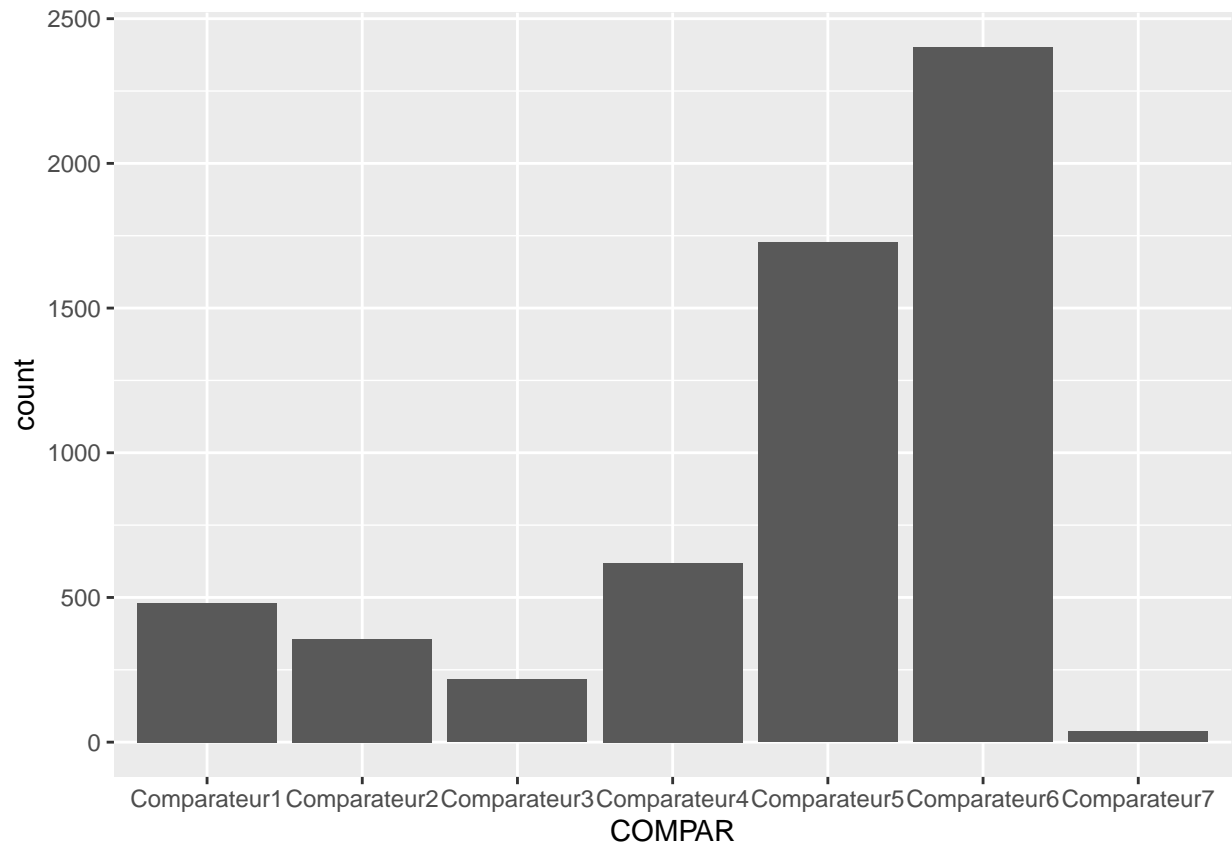
## Visualisation de nos variables

```
barplot(table(contra_comp$COMPAR))
```





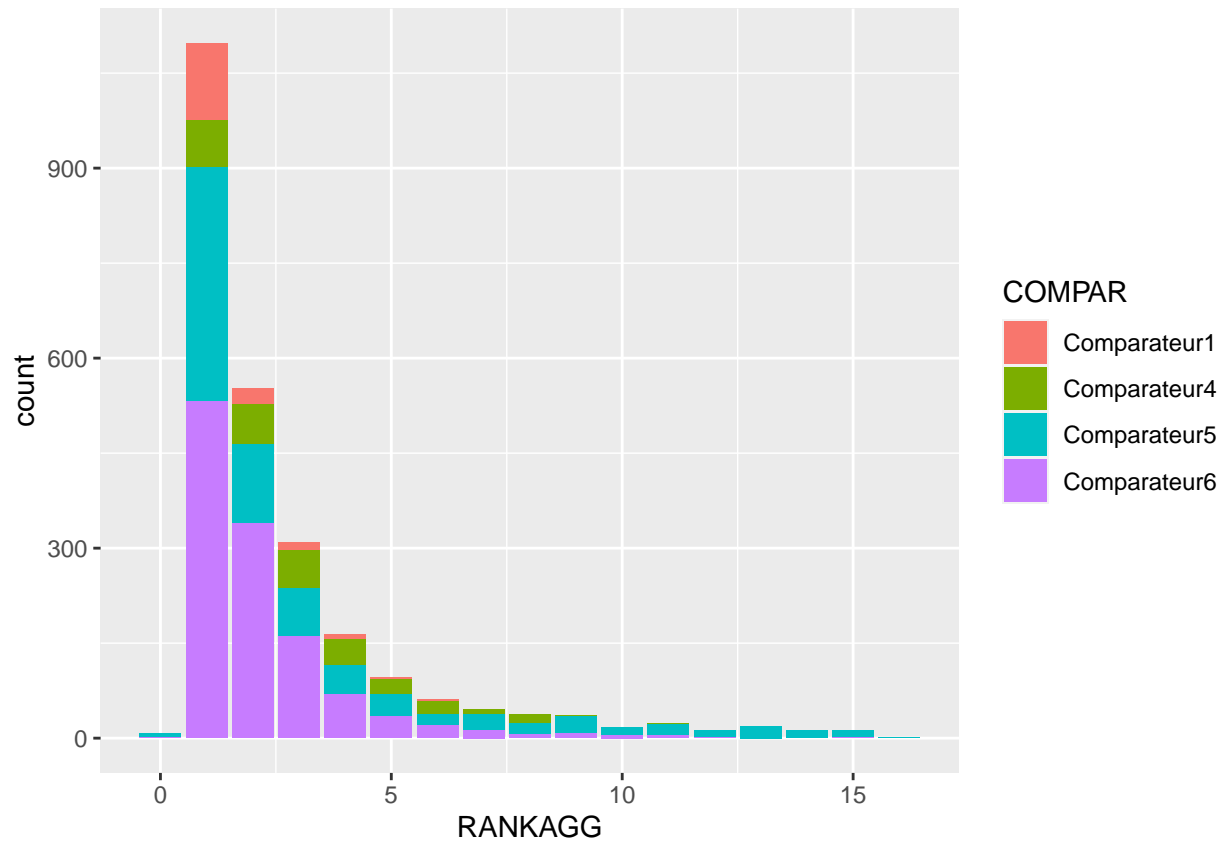
```
ggplot()+geom_bar(data=contra_comp,aes(x=COMPAR))
```



Le comparateur 6 et le comparateur5 sont les plus utilisés. Il serait intéressant de vérifier notre rang sur chacun de ses comparateurs.

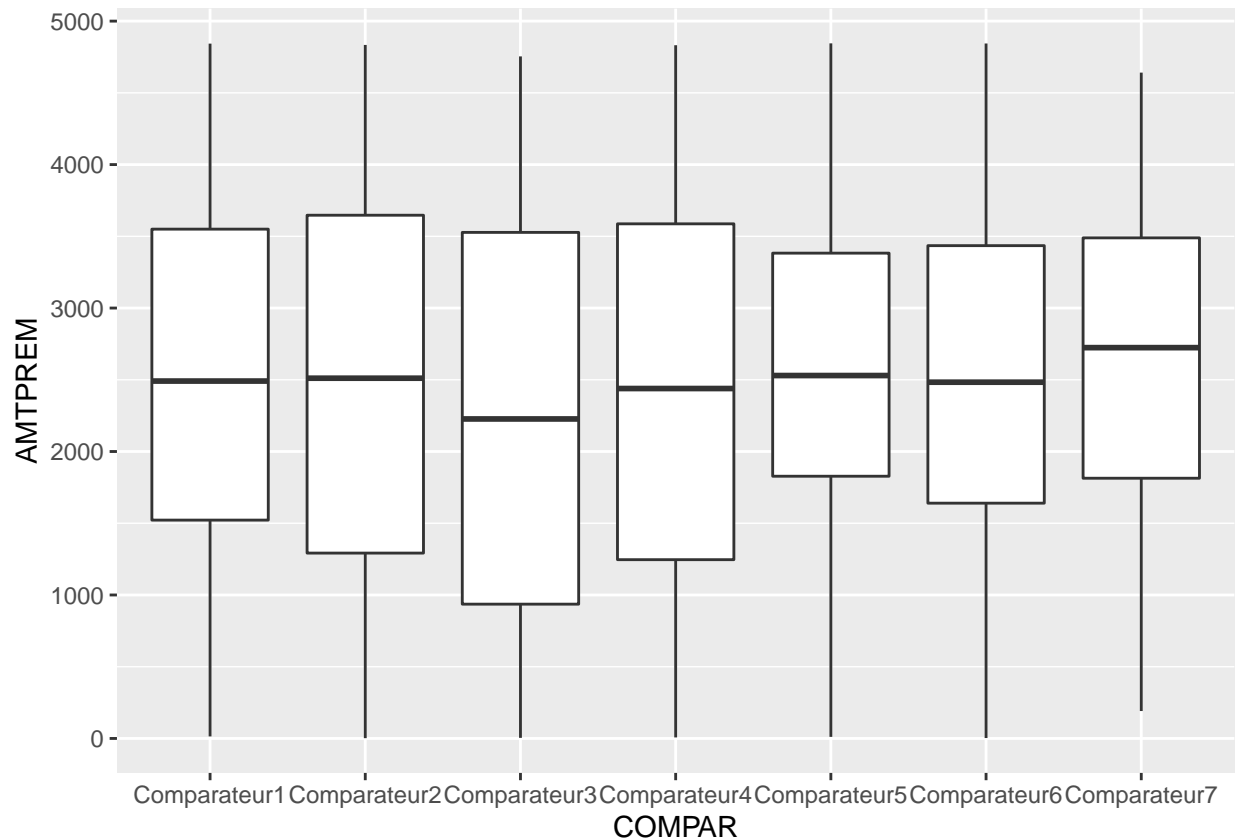
```
ggplot()+geom_bar(data=contra_comp,aes(x=RANKAGG,fill=COMPAR))
```

```
## Warning: Removed 3328 rows containing non-finite values (stat_count).
```



On voit clairement que notre assureur est mieux classé sur les “comparateur6” et “comparateur5” suivi du “comparateur4”. On peut donc conclure que mieux notre assureur est bien classé et plus les assurés souscrivent à nos polices d’assurance.

```
ggplot(contra_comp, aes(x=COMPAR,y=AMTPREM)) +
  geom_boxplot()
```



#### Interprétation :

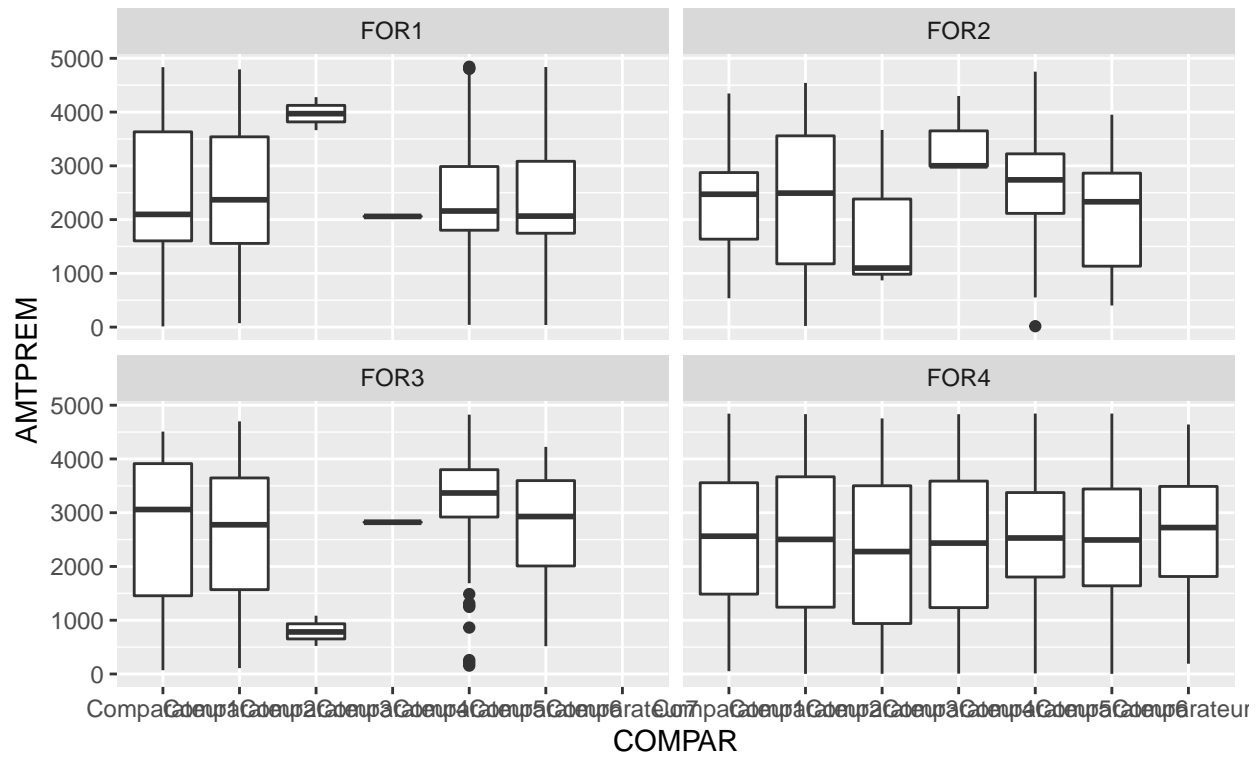
- On comprend mieux les raisons pour lesquelles les clients ont plus souscrits via les comparateur6 et comparateur5. Car ces primes sur ces comparateurs sont plus faibles que tous les autres.
- Via Le comparateur7, les clients bénéficient de primes à bas prix, on pourrait se demander pourquoi après les comparateur6 et comparateur5, ce n'est pas le comparateur7 qui suit, mais plutôt le 4, où pourtant les primes sont assez élevées.
- Un piste de solution serait dû au fait que notre assureur est bien classé sur les comparateurs **6, 5, 4 et 1**. Ce qui n'est pas le cas du comparateur7 où nous sommes mal classés( confert diagramme du rang en fonction des comparateurs).

**Conclusion :**On peut donc affirmer que le classement sur le comparateur est d'une importance capitale.

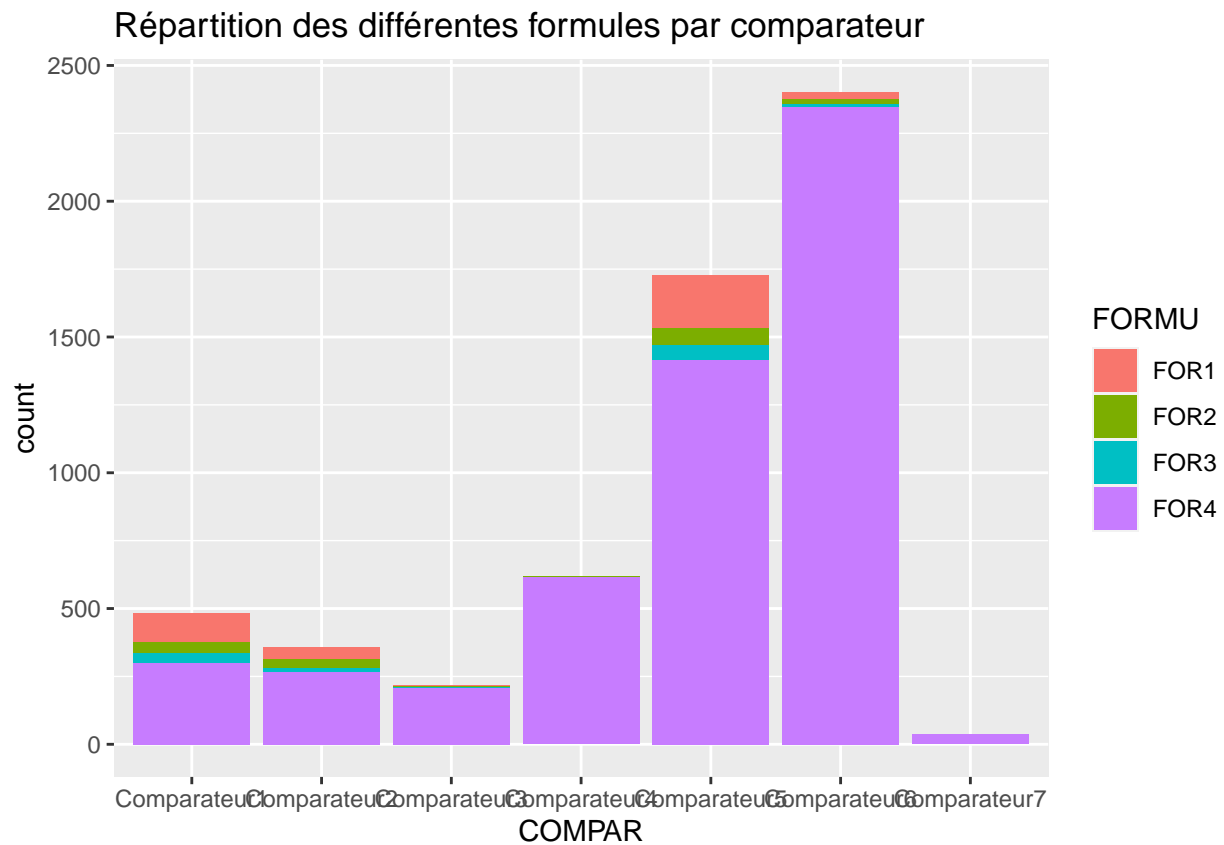
#### Analyse des primes en fonction des comparateurs et des différentes formules souscrites

```
ggplot(contra_comp, aes(x=COMPAR,y=AMTPREM))+
geom_boxplot()+facet_wrap(~FORMU)+ggtitle("Analyse des primes en fonction des comparateurs \n et des di.
```

## Analyse des primes en fonction des comparateurs et des différentes formules des contrats



```
ggplot(contra_comp, aes(x=COMPAR, fill=FORMU))+geom_bar()+ggtitle("Répartition des différentes formules")
```



```
#install.packages("flexdashboard")

#install.packages("htmltools")

#install.packages("shiny")

library(flexdashboard)

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

library(htmltools)

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

library(shiny)

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

ui <- fluidPage(

  fillCol(height = 600, flex = c(NA, 1),

  inputPanel(

    selectInput("COMPAR", "Comparateur:", choices =levels(contra$COMPAR) )

  ),
```

```

    plotOutput("formuPlot", height = "100%")
  )
)

server <- function(input, output) {

  output$formuPlot <- renderPlot({

    ggplot(subset(contra_comp, COMPAR==input$COMPAR), aes(x=FORMU)) + geom_bar()

  })

}

shinyApp(ui=ui, server = server)

```

## PhantomJS not found. You can install it with `webshot::install_phantomjs()`. If it is installed, please

Shiny applications not supported in static R Markdown documents

### Interprétation :

- La formule 4 est la formule la plus souscrite et ce notamment via le comparateur 6.
- De même, on constate que le comparateur6 et comparateur5 proposent des primes plus basses pour les formules d'assurance. Ce qui explique une fois de plus que les assurés ont majoritairement souscrit à des polices via les comparateurs 6 et 7.
- Le comparateur propose des formules trop chères.

---



---



---



---

## Analyse des caractéristiques des clients ayant souscrit via les différents comparateurs

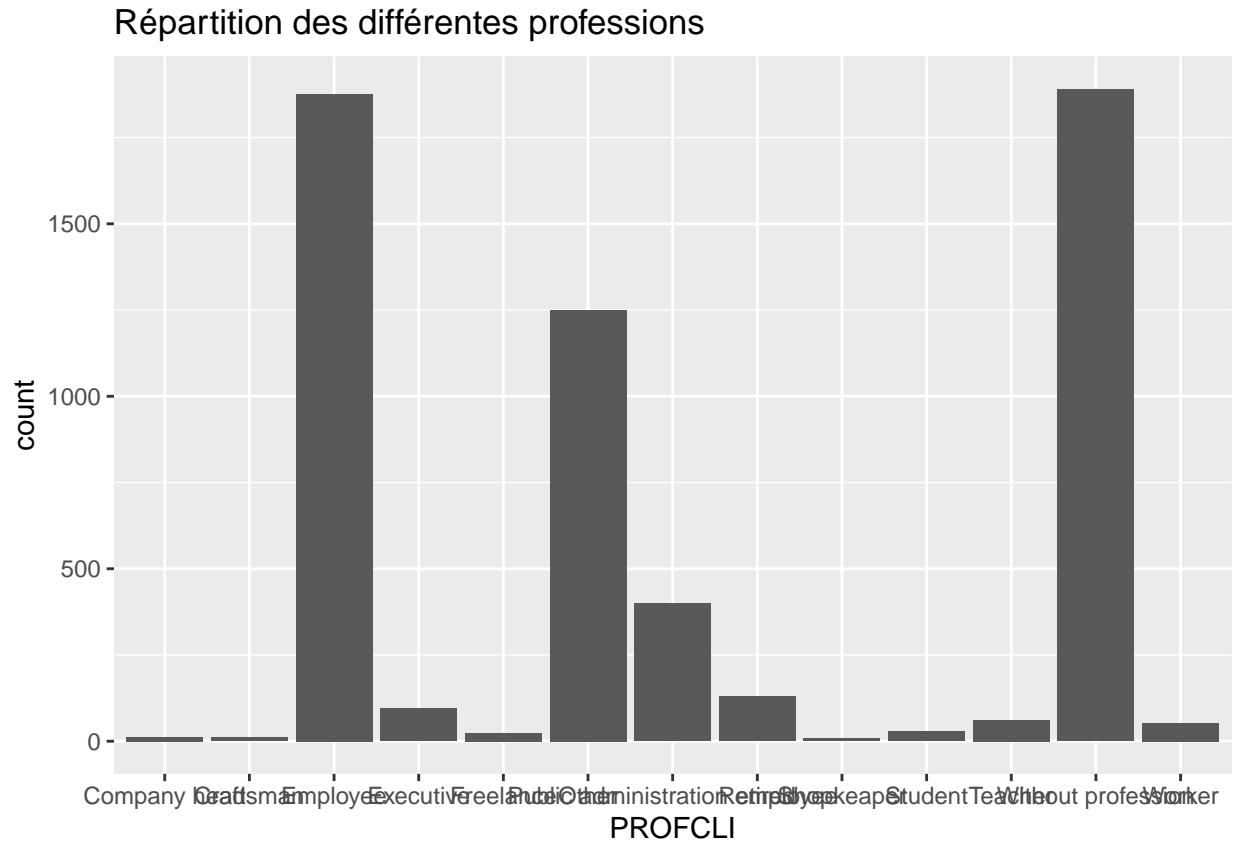
1. la situation professionnelle

```
table(contra_comp[,PROFCLI], useNA = "always")
```

```
##
##               Company head               Craftsman
##                12                11
##             Employee             Executive
##            1876                95
##             Freelance                Other
##                24            1250
## Public administration employee        Retired
##                399            129
##                Shopkeeper        Student
```

```
##          7          28
##      Teacher      Without profession
##      62          1890
##      Worker      <NA>
##      53          0
```

```
ggplot()+geom_bar(data=contra_comp,aes(x=PROFCLI))+ggtitle("Répartition des différentes professions")
```

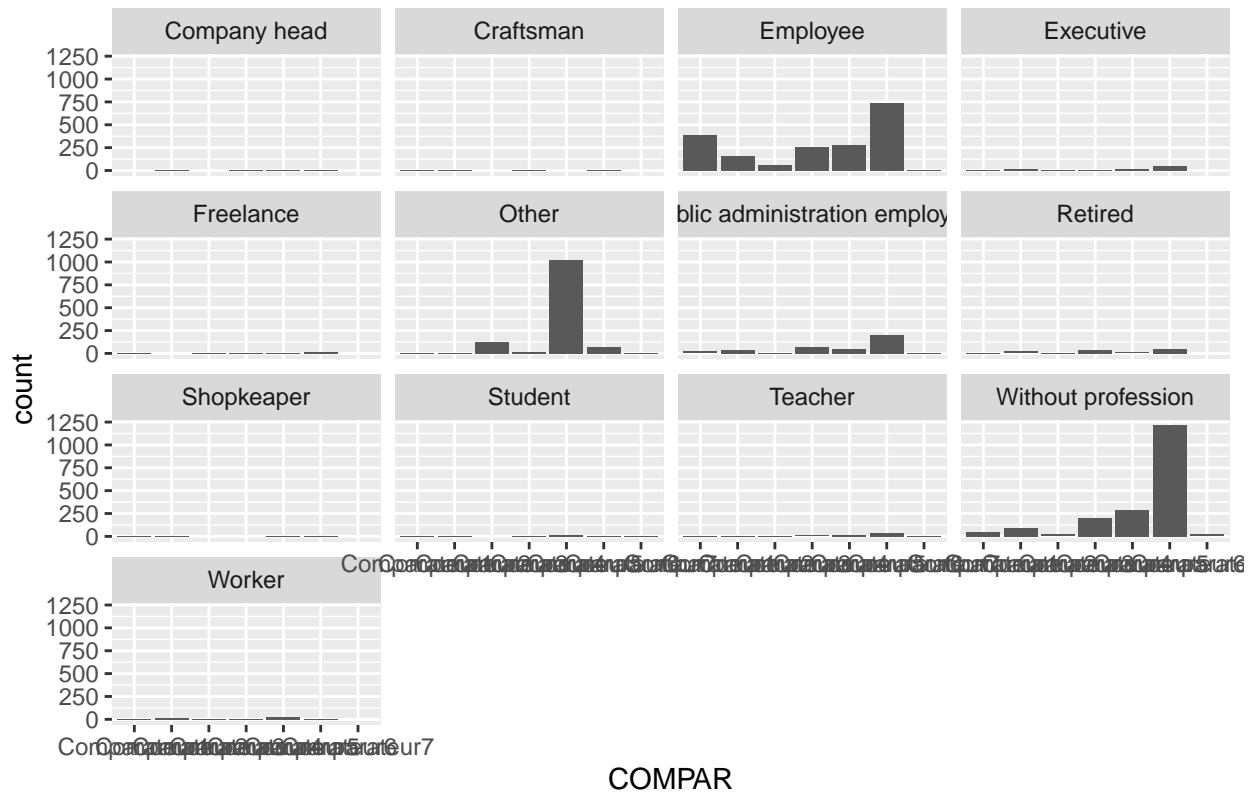


Remarque : Notre base des assurés est constituée majoritairement de personnes sans emploi et d'employés.

```
ggplot(contra_comp, aes(x=COMPAR))+geom_bar()+facet_wrap(~PROFCLI)+ggtitle("Répartition des comparateurs")
```



## Répartition des comparateurs par profession



### Interprétation :

- On remarque qu'indépendamment de la profession, le comparateur6 reste le comparateur qui attire le plus de clients.
- Il est important de remarquer que les clients qui sont d'autres professions("other") ont pour la plupart souscrits via le comparateur5.

### 2. le sexe

```
tabsex=table(contra_comp$SEXCLI,useNA = "always")
```

```
print(tabsex)
```

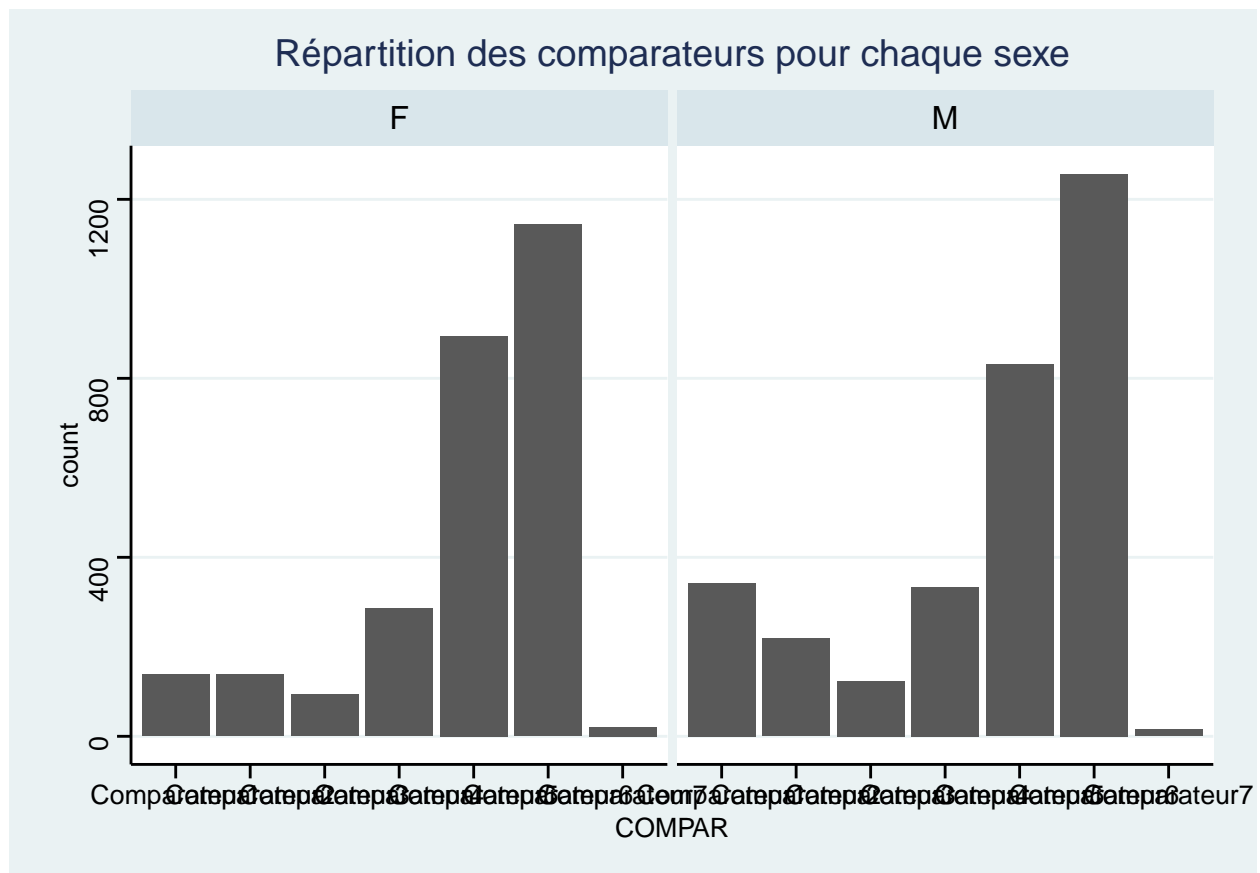
```
##
##      F      M <NA>
## 2715 3121      0
```

```
tabsex_freq=tabsex*100/nrow(contra_comp)
```

```
tabsex_freq
```

```
##
##      F      M      <NA>
## 46.52159 53.47841 0.00000
```

```
ggplot(contra_comp, aes(x=COMPACT))+geom_bar()+facet_wrap(~SEXCLI)+ggtitle("Répartition des comparateurs
```



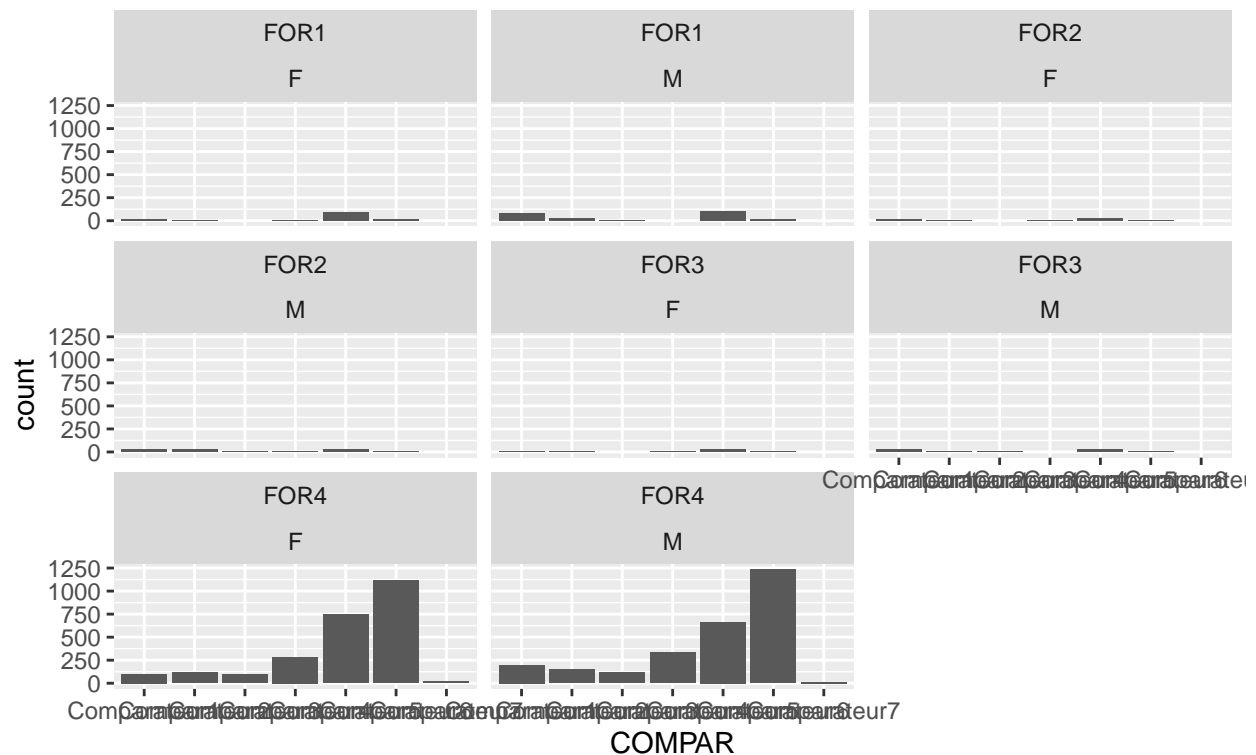
Interprétation:

- On peut dire que le sexe des différents clients ne permet pas d'expliquer le choix des différents comparateurs.
- Les clients hommes et femmes ne sont pas si différents dans leurs choix. Ils ont souscrit tous majoritairement via le comparateur 6 et 7.

```
library("ggsci")
```

```
ggplot(contra_comp, aes(x=COMPACTEUR))+geom_bar()+facet_wrap(FORMULE~SEXCLIENT)+ggtitle("Répartition du choix des comparateurs par sexe")
```

## Répartition du choix des différents comparateurs pour chaque formule et par sexe



### Interprétation:

Ce graphique confirme le fait que les clients hommes et femmes ont sensiblement les mêmes caractéristiques. Ils ont tous majoritairement souscrit à la **Formule 4** et cela via les **comparateurs 6 et 7**.

### 3. Le statut marital

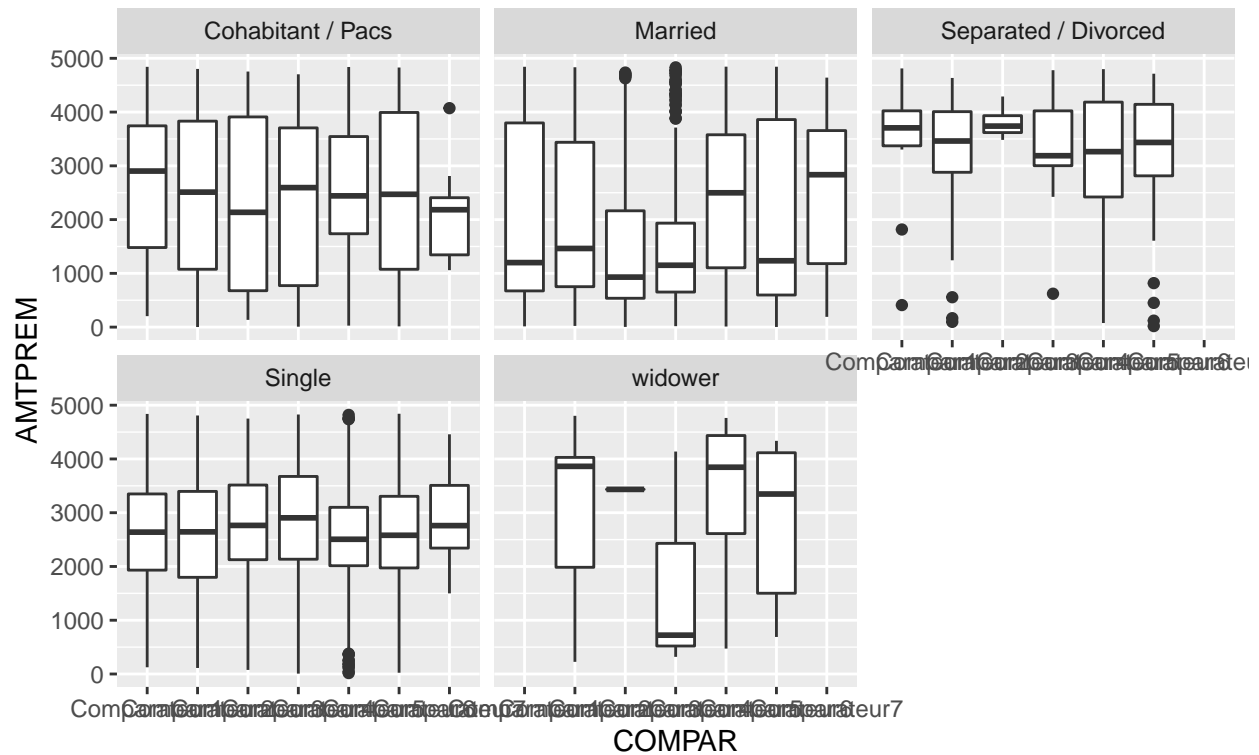
```
table(contra$MARCLI)
```

```
##
## Cohabitant / Pacs Married Separated / Divorced
##           794      2138           255
##           Single   widower
##           3776       30
```

```
ggplot(contra_comp, aes(x=COMPAR, y=AMTPREM)) +
```

```
geom_boxplot() + facet_wrap(~MARCLI) + ggtitle("Analyse des primes en fonction des comparateurs \n et du statut marital")
```

## Analyse des primes en fonction des comparateurs et du statut marital du client



```
tab_prim=contra_comp[,.(moyenne_prim=mean(AMTPREM)),by=MARCLI]
```

```
print(tab_prim)
```

```
##           MARCLI  moyenne_prim
## 1:      Single      2652.588
## 2:      Married      2121.465
## 3: Cohabitant / Pacs      2473.952
## 4: Separated / Divorced      3167.114
## 5:      widower      2902.320
```

```
tab_compa=contra_comp[,.(moyenne_prim=mean(AMTPREM)),by=COMPAR]
```

```
print(tab_compa)
```

```
##           COMPAR  moyenne_prim
## 1: Comparateur1      2493.732
## 2: Comparateur2      2445.583
## 3: Comparateur3      2290.065
## 4: Comparateur4      2427.729
## 5: Comparateur5      2544.475
## 6: Comparateur6      2479.122
## 7: Comparateur7      2635.944
```

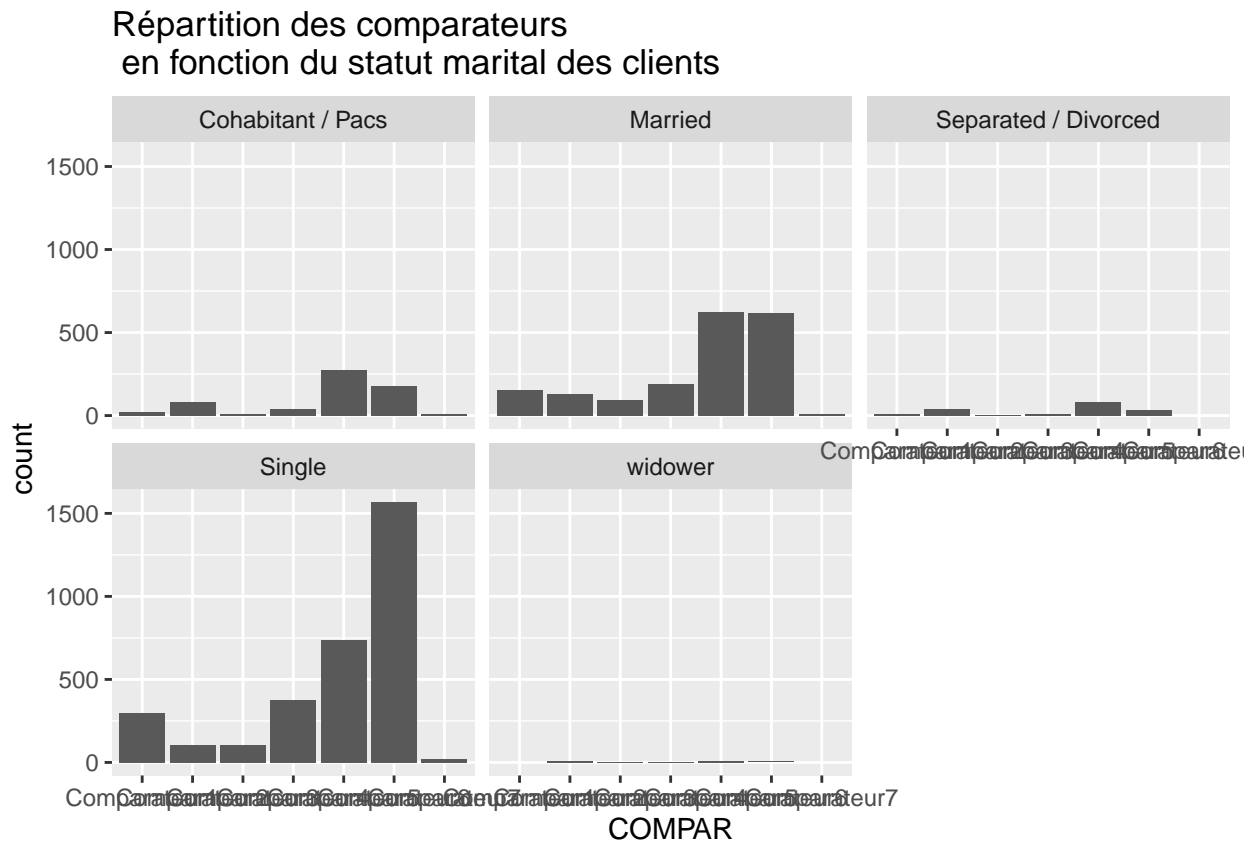
### Interprétations:

Des schémas ci-dessus, on peut dire que :

- Les clients mariés paient des primes plus élevées que tous les autres suivi des veufs.

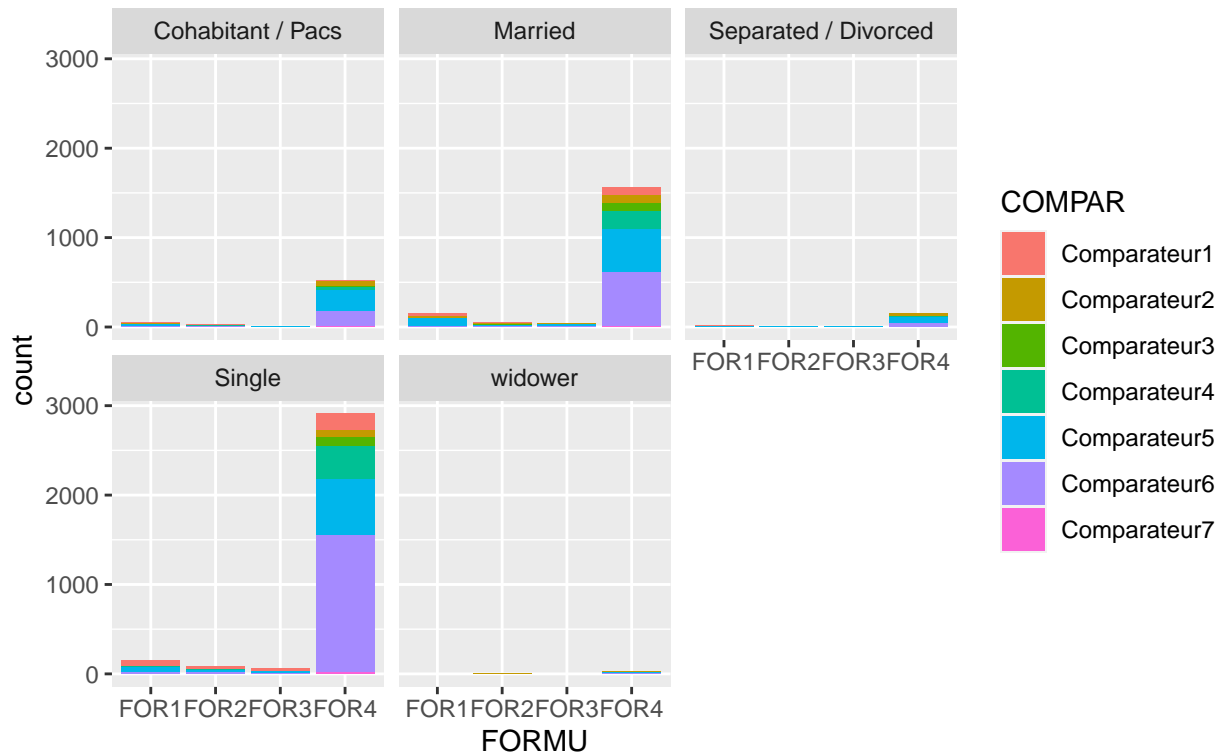
- les célibataires paient les primes les moins élevées.
- Les clients ayant souscrit un contrat via le comparateur 4, ont payé des primes plus élevées.
- Les mariés qui ont souscrit via le comparateur 5 paient les primes les plus basses parmi les clients mariés( il serait intéressant d'analyser plus tard les formules les plus appréciées et si le nombre d'enfants a eu une influence dans le choix des différentes formules).

```
ggplot(contra_comp, aes(x=COMPAR))+
geom_bar()+facet_wrap(~MARCLI)+ggtitle("Répartition des comparateurs \n en fonction du statut marital d
```



```
ggplot(contra_comp, aes(x=FORMU,fill=COMPAR))+
geom_bar()+facet_wrap(~MARCLI)+ggtitle("Répartition des formules et du comparateur \n pour chaque statut
```

## Répartition des formules et du comparateur pour chaque statut marital des clients



### Remarque:

- La formule 4 est très appréciée qu'importe le statut marital du client.
- La formule 4 est très appréciée notamment des célibataires et des mariés.
- Qu'importe le statut marital, les clients ont souscrit à la formule majoritairement via les **comparateurs 6 et 5**. Cela s'explique notamment par le fait que les primes souscrites via les comparateurs 6 et 5, sont les plus basses de notre portefeuille; et ces comparateurs sont aussi ceux où notre assureur est le mieux classé.

**On voit à clairement un rapport entre le classement de notre assureur sur un comparateur et le nombre d'assurés ayant souscrit via ce comparateur.**

- Analyse des choix des formules d'assurance par statut marital et pour chaque comparateur:

```
ggplot(contra_comp, aes(x=FORMU))+
geom_bar()+facet_wrap(COMPAR~MARCLI)+ggtitle("Répartition des formules \n et par comparateur et statut marital")
```

## Répartition des formules et par comparateur et statut marital des clients



Remarque :

Qu'importe le comparateur et le statut marital, les assurés ont plus souscrit à la formule 4, elle reste la plus avantageuse pour nos assurés.

- Moyenne des primes par formule

```
tab_form=contra_comp[,.(moy_prim=mean(AMTPREM),nombre=.N),by=.(FORMU)]
print(tab_form)
```

```
## FORMU moy_prim nombre
## 1: FOR1 2409.190 368
## 2: FOR4 2484.656 5186
## 3: FOR2 2458.055 164
## 4: FOR3 2829.466 118
```

### Interprétations:

- le tableau ci dessus,explique mieux pourquoi la formule 4 est la plus appréciée des clients, elle offre les primes les plus basses.

#### 4. Le nombre d'enfants

```
nbre_enfant=as.integer(!is.na(contra_comp$IDCHI1))+as.integer(!is.na(contra_comp$IDCHI2))+as.integer(!is.na(contra_comp$IDCHI3))
contra_comp[,nombre_enfants:=nbre_enfant]
```

- Statistiques sur le nombre d'enfants

```
summary(contra_comp$nombre_enfants)
```

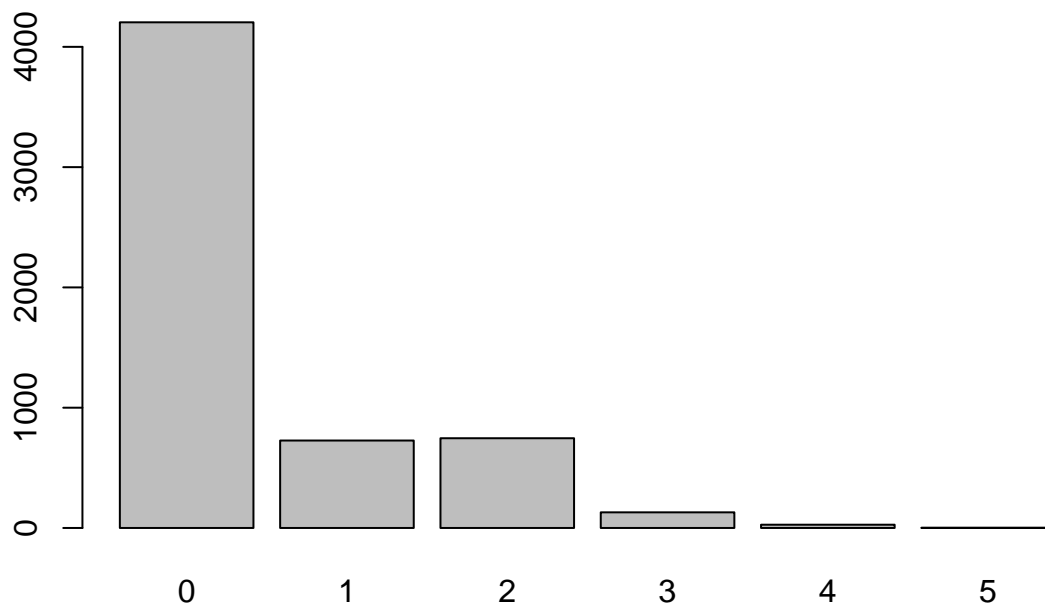
```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0000 0.0000 0.0000 0.4673 1.0000 5.0000
```

```
tab_enf=table(contra_comp$nombre_enfants)
```

```
print(tab_enf)
```

```
##
##      0      1      2      3      4      5
## 4204  727  746  130   27    2
```

```
barplot(table(contra_comp$nombre_enfants))
```



```
tab_freq=tab_enf*100/nrow(contra_comp)
```

```
print(tab_freq)
```

```
##
##      0      1      2      3      4      5
## 72.03564085 12.45716244 12.78272790 2.22755312 0.46264565 0.03427005
```

Remarque:

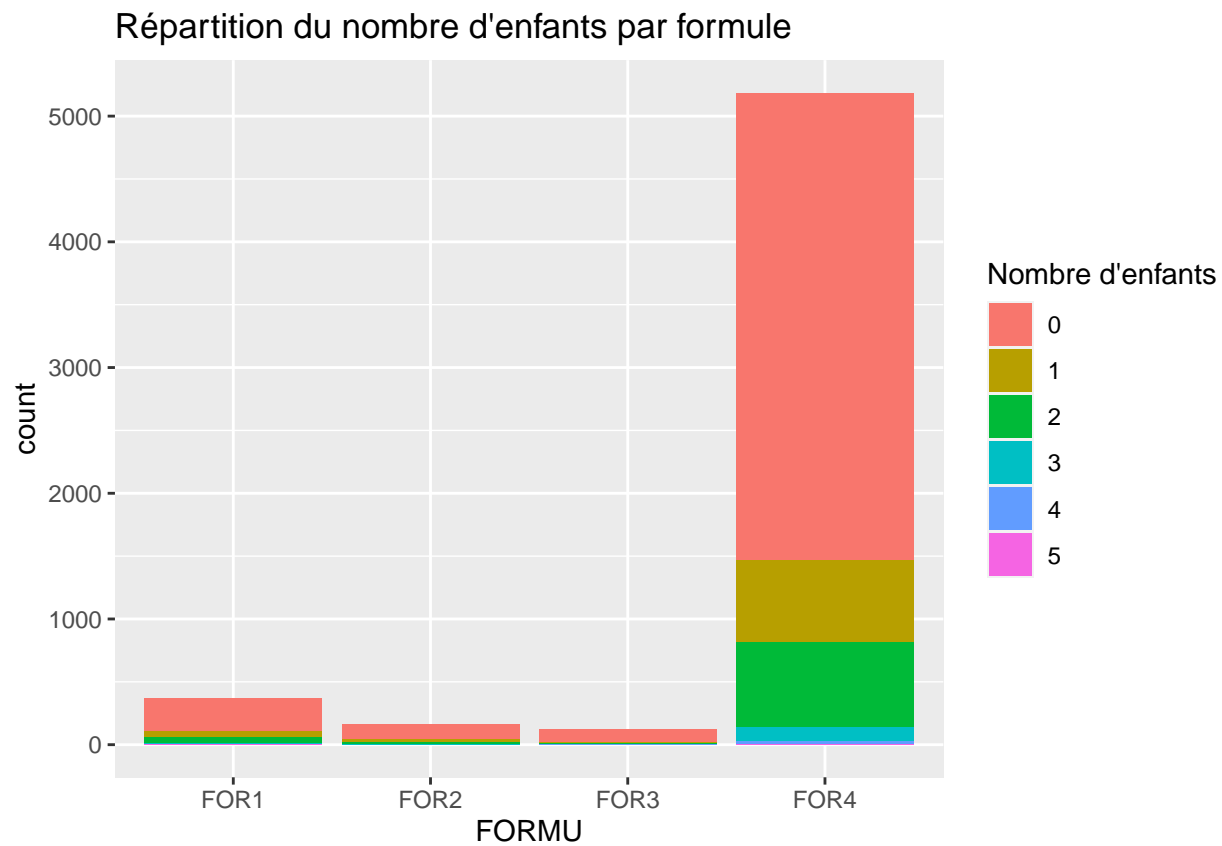
Notre base est constituée à plus de 70% par des clients n'ayant aucun enfant et à plus de 95% de clients ayant moins de 2 enfants.

- Nous allons donc analyser le choix des différentes formules et des comparateurs en fonction du nombre



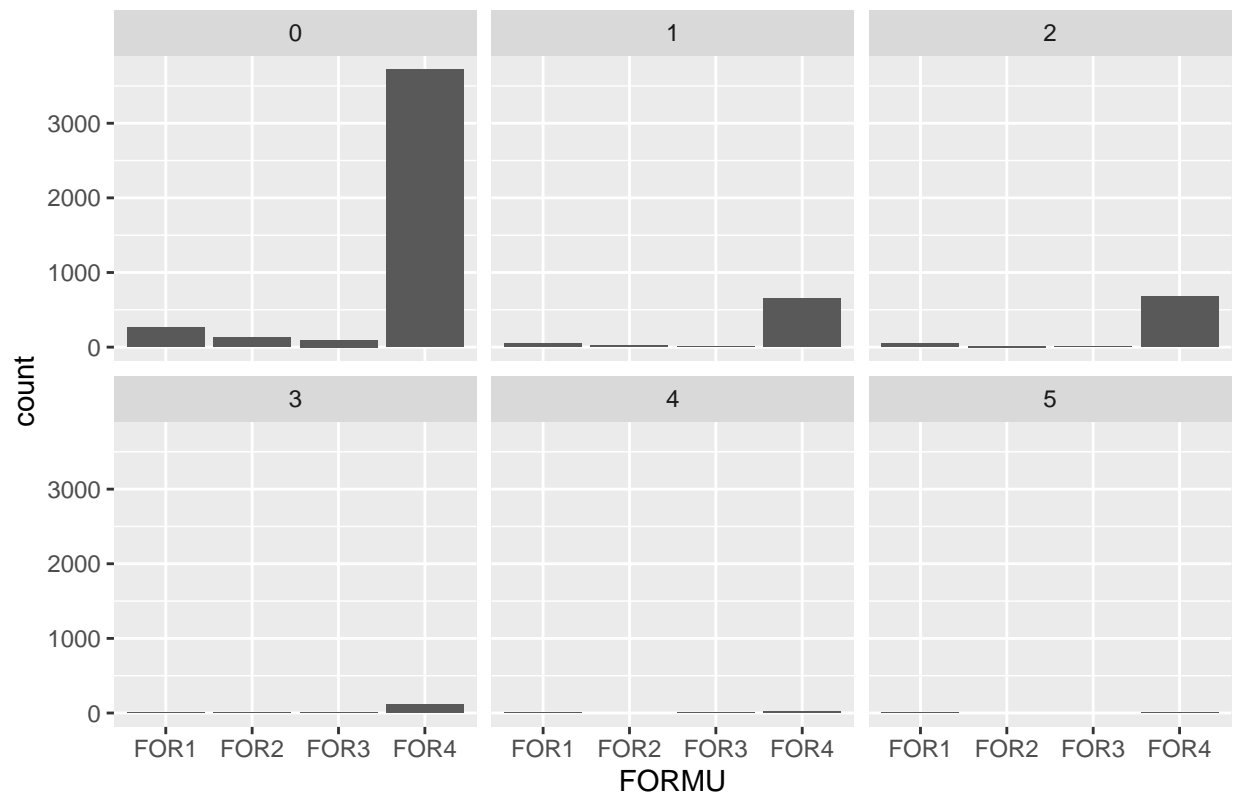
d'enfants de chaque client, pour ainsi voir si le nombre d'enfants influence le choix de la formule.

```
ggplot(data=contra_comp,aes(x=FORMU,fill=as.factor(contra_comp$nombre_enfants)))+geom_bar()+labs(fill="")
```



```
ggplot(data=contra_comp,aes(x=FORMU))+geom_bar()+ggtitle("Analyse du choix des formules en fonction du nombre d'enfants")
```

## Analyse du choix des formules en fonction du nombre d'enfants

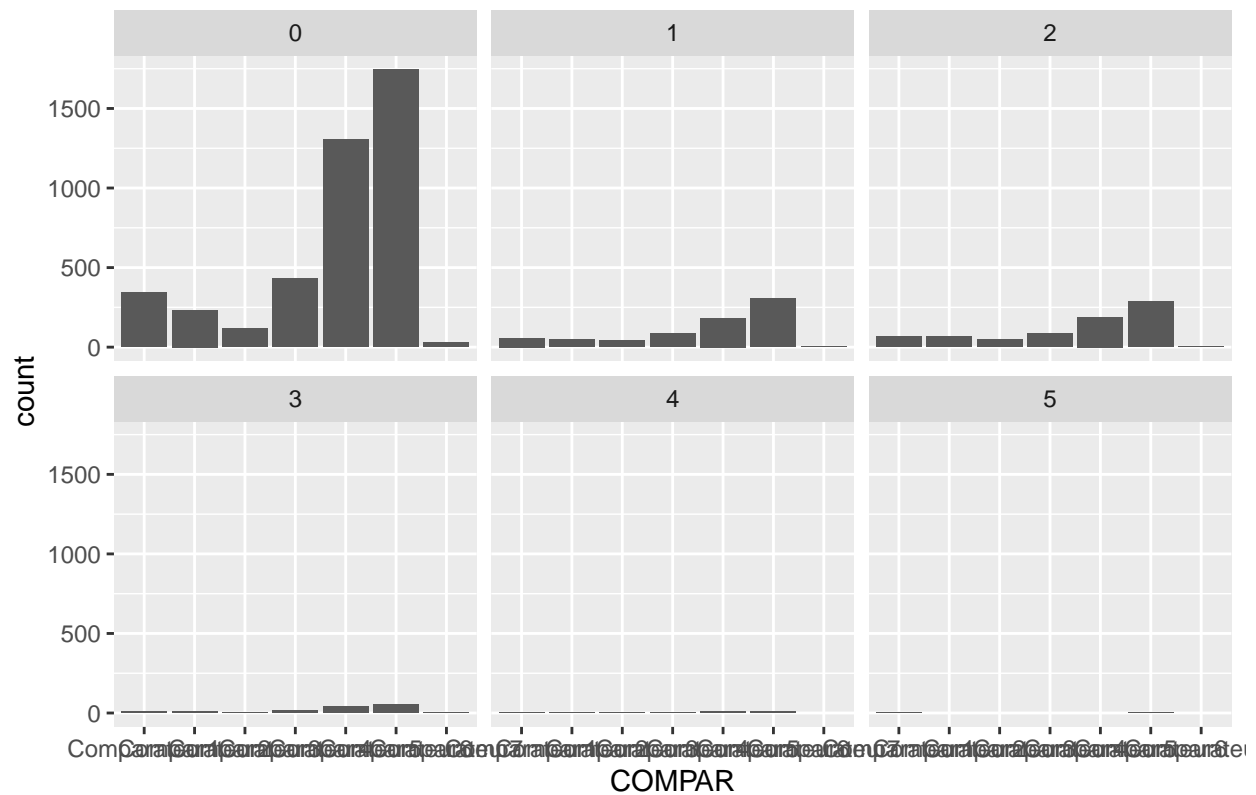


### Interprétation :

Qu'importe le nombre d'enfants, la formule 4 demeure la plus appréciée. *texte en italique*

```
ggplot(data=contra_comp,aes(x=COMPAR))+geom_bar()+ggtitle("Analyse des comparateurs en fonction du nomb
```

## Analyse des comparateurs en fonction du nombre d'enfants



### Interprétation:

Le nombre d'enfants des clients n'a pas influencé le fait que la plus grande partie de nos assurés aient souscrit via les comparateurs 6 et 5.

#### 5. Durée des contrats

```
table(is.na(contra_comp$DATEEFF))
```

```
##
## FALSE
## 5836
```

Tous les contrats possèdent une date d'effet de garantie.

```
table(is.na(contra_comp$DATECAN))
```

```
##
## FALSE  TRUE
## 1149  4687
```

Il y'a 4687 contrats sans date de résiliation et 1149 contrats résiliés.

```
contra_comp[DATECAN=="NA",resilie:="Non"]
```

```
contra_comp[is.na(DATECAN),resilie:="Non"]
```

```
contra_comp[!is.na(DATECAN),resilie:="Oui"]
```

```
table(contra_comp$resilie)
```

```
##
##   Non   Oui
## 4687 1149
```

- Création de la variable représentant la durée des contrats pour les contrats qui ont été résiliés:

```
#clients ayant résilié leurs contrats
```

```
contra_resil=contra_comp[resilie=="Oui"]
```

```
print(contra_resil)
```

```
##          IDPOL    DATEEFF      IDCLI    IDPART    IDCHI1    IDCHI2 IDCHI3
##   1: 2C7EAEF380 2011-02-10 7A61ADDB61      <NA>      <NA>      <NA> <NA>
##   2: F1125A72F9 2011-07-10 4E5D3AA63C 32D5615B20 2EF179B6C7 B09D69E316 <NA>
##   3: 0F5BF8EA40 2011-08-10 283B50C751      <NA>      <NA>      <NA> <NA>
##   4: 73FB728155 2011-08-10 2DF3A65E55      <NA>      <NA>      <NA> <NA>
##   5: 21BFCE62CB 2012-01-01 9C968CCD1F      <NA>      <NA>      <NA> <NA>
## ---
## 1145: CD2B50E2C8 2015-01-01 0459108AB7      <NA>      <NA>      <NA> <NA>
## 1146: 75B344BC5B 2015-01-01 7BD50DF5BD 8C1E9ADB9E E3EBF67A08 OB47EABB20 <NA>
## 1147: 393975A549 2015-01-01 4C7F9A4B88      <NA>      <NA>      <NA> <NA>
## 1148: DA3E2F195B 2015-01-01 E8E1D8AB93      <NA>      <NA>      <NA> <NA>
## 1149: A240758C99 2015-01-01 66D9A5E612      <NA>      <NA>      <NA> <NA>
##          IDCHI4 IDCHI5 IDCHI6 FORMU    DATECAN
##   1:   <NA>   <NA>    NA  FOR1 2017-12-31
##   2:   <NA>   <NA>    NA  FOR4 2016-12-31
##   3:   <NA>   <NA>    NA  FOR2 2016-12-31
##   4:   <NA>   <NA>    NA  FOR3 2016-12-31
##   5:   <NA>   <NA>    NA  FOR4 2016-12-31
## ---
## 1145:   <NA>   <NA>    NA  FOR4 2015-01-13
## 1146:   <NA>   <NA>    NA  FOR1 2014-01-13
## 1147:   <NA>   <NA>    NA  FOR4 2014-01-13
## 1148:   <NA>   <NA>    NA  FOR2 2013-01-13
## 1149:   <NA>   <NA>    NA  FOR4 2013-01-13
##
##                                REACAN AMTPREM RANKAGG      COMPAR
##   1:                                Non-payment of the premium 1455      1 Comparateur1
##   2:                                Non-payment of the premium 4242      4 Comparateur1
##   3:                                Non-payment of the premium 1977      1 Comparateur1
##   4:                                Non-payment of the premium 3628      2 Comparateur1
##   5:                                Non-payment of the premium 4738      2 Comparateur1
## ---
## 1145:                                Non-payment of the premium 3680      NA Comparateur1
## 1146:                                Cancel renewal policy 4794      NA Comparateur2
## 1147:                                Cancel renewal policy 1631      NA Comparateur2
## 1148:                                Non-payment of the premium 1801      NA Comparateur2
## 1149: Cancellation on renewal from the customer 3802      NA Comparateur4
##
##          MARCLI SEXCLI    DTNCLI          PROFCLI    CP SEXCJ
##   1:          Single      M 1980-12-16      Employee 29200 <NA>
##   2:          Married      M 1968-12-22      Employee 1700    F
##   3:          Single      M 1977-06-08      Employee 45260 <NA>
##   4:          Single      M 1960-09-05      Employee 74130 <NA>
```

```
##      5:          Single      F 1951-02-19          Retired 14600 <NA>
##      ---
## 1145:          Single      M 1955-01-13          Retired 95220 <NA>
## 1146: Cohabitant / Pacs      F 1969-07-16          Employee 29470      M
## 1147:          Single      M 1990-12-01 Without profession 13011 <NA>
## 1148: Separated / Divorced      M 1961-09-17 Without profession 67200 <NA>
## 1149:          Single      F 1968-04-05          Employee 34560 <NA>
##          DTNCJ          PROFCLJ nombre_enfants resilie
##      1:          <NA>          <NA>          0      Oui
##      2: 1978-11-15          Employee          2      Oui
##      3:          <NA>          <NA>          0      Oui
##      4:          <NA>          <NA>          0      Oui
##      5:          <NA>          <NA>          0      Oui
##      ---
## 1145:          <NA>          <NA>          0      Oui
## 1146: 1966-11-28 Without profession          2      Oui
## 1147:          <NA>          <NA>          0      Oui
## 1148:          <NA>          <NA>          0      Oui
## 1149:          <NA>          <NA>          0      Oui
```

*#conversion de caractères en dates*

```
date_deb=as.Date(contra_resil$DATEEFF,format = "%Y-%m-%d")
```

```
date_fin=as.Date(contra_resil$DATECAN,format = "%Y-%m-%d")
```

```
duree=date_fin-date_deb
```

```
head(duree)
```

```
## Time differences in days
```

```
## [1] 2516 2001 1970 1970 1826 1848
```

```
contra_resil[,dureeContra:=duree]
```

```
head(contra_resil)
```

```
##          IDPOL      DATEEFF      IDCLI      IDPART      IDCHI1      IDCHI2 IDCHI3
## 1: 2C7EAEF380 2011-02-10 7A61ADDB61      <NA>      <NA>      <NA>      <NA>
## 2: F1125A72F9 2011-07-10 4E5D3AA63C 32D5615B20 2EF179B6C7 B09D69E316 <NA>
## 3: 0F5BF8EA40 2011-08-10 283B50C751      <NA>      <NA>      <NA>      <NA>
## 4: 73FB728155 2011-08-10 2DF3A65E55      <NA>      <NA>      <NA>      <NA>
## 5: 21BFCE62CB 2012-01-01 9C968CCD1F      <NA>      <NA>      <NA>      <NA>
## 6: 0EF51BD7B0 2011-12-10 51305185AA      <NA>      <NA>      <NA>      <NA>
##          IDCHI4 IDCHI5 IDCHI6 FORMU      DATECAN          REACAN AMTPREM
## 1: <NA> <NA>      NA FOR1 2017-12-31 Non-payment of the premium 1455
## 2: <NA> <NA>      NA FOR4 2016-12-31 Non-payment of the premium 4242
## 3: <NA> <NA>      NA FOR2 2016-12-31 Non-payment of the premium 1977
## 4: <NA> <NA>      NA FOR3 2016-12-31 Non-payment of the premium 3628
## 5: <NA> <NA>      NA FOR4 2016-12-31 Non-payment of the premium 4738
## 6: <NA> <NA>      NA FOR4 2016-12-31 Non-payment of the premium 4320
##          RANKAGG      COMPAR MARCLI SEXCLI      DTNCLI PROFCLI      CP SEXCJ
## 1:      1 Comparateur1 Single      M 1980-12-16 Employee 29200 <NA>
## 2:      4 Comparateur1 Married      M 1968-12-22 Employee 1700      F
## 3:      1 Comparateur1 Single      M 1977-06-08 Employee 45260 <NA>
## 4:      2 Comparateur1 Single      M 1960-09-05 Employee 74130 <NA>
```

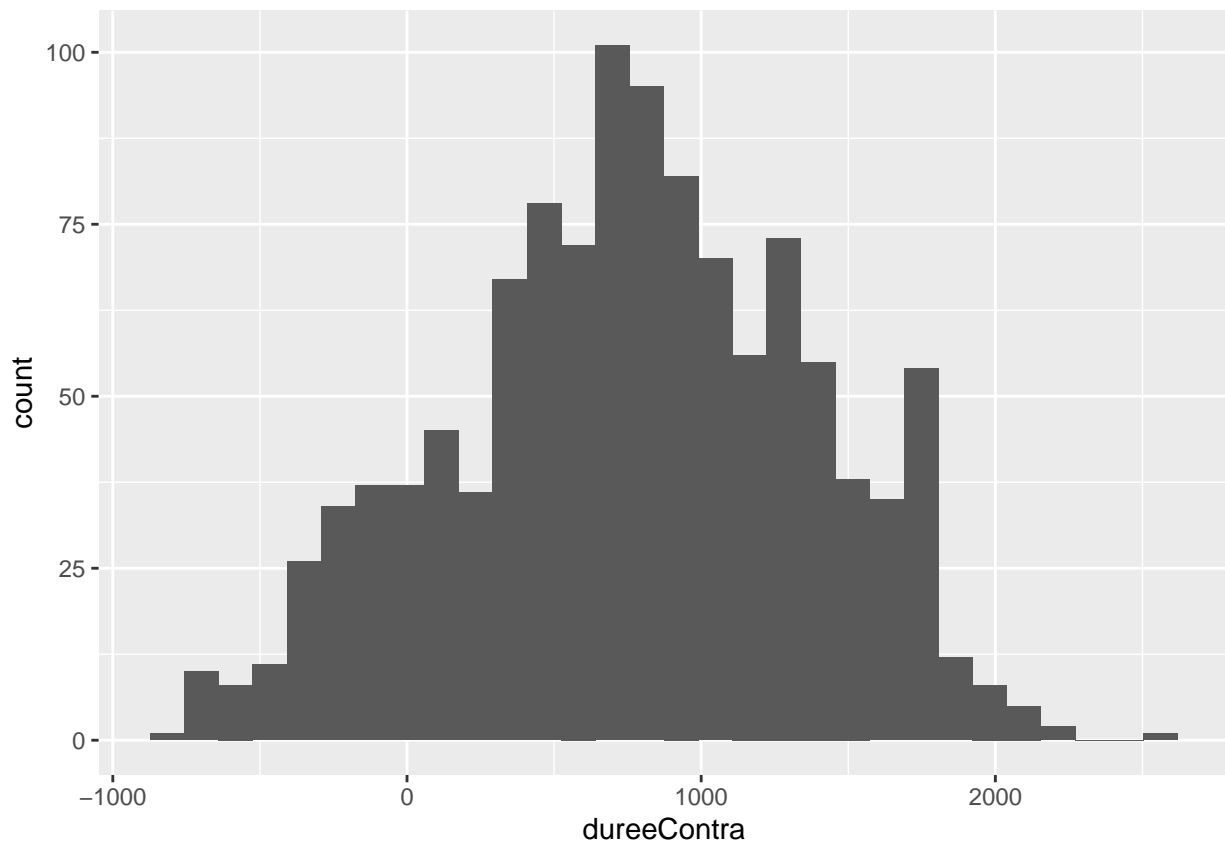
```
## 5:      2 Comparateur1 Single      F 1951-02-19 Retired 14600 <NA>
## 6:      6 Comparateur1 Single      F 1962-07-12 Employee 66600 <NA>
##      DTNCJ  PROF CJ nombre_enfants resilie dureeContra
## 1:      <NA>      <NA>           0      Oui  2516 days
## 2: 1978-11-15 Employee           2      Oui  2001 days
## 3:      <NA>      <NA>           0      Oui  1970 days
## 4:      <NA>      <NA>           0      Oui  1970 days
## 5:      <NA>      <NA>           0      Oui  1826 days
## 6:      <NA>      <NA>           0      Oui  1848 days
```

```
#distribution de la durée des contrats
```

```
ggplot()+geom_histogram(data=contra_resil,aes(x = dureeContra))
```

```
## Don't know how to automatically pick scale for object of type difftime. Defaulting to continuous.
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



Remarque:

Il y a des dates négatives, qui ne sont pas conformes. Il faut les retirer de la base de données.

```
#les contrats non conformes
```

```
a=contra_resil[which(dureeContra<0)]
```

```
head(a[,dureeContra])
```

```
## Time differences in days
```

```
## [1] -320 -144 -46 -46 -47 -366
```

```
contra_resil_clean=contra_resil[which(dureeContra>=0)]
```

```
head(contra_resil_clean)
```

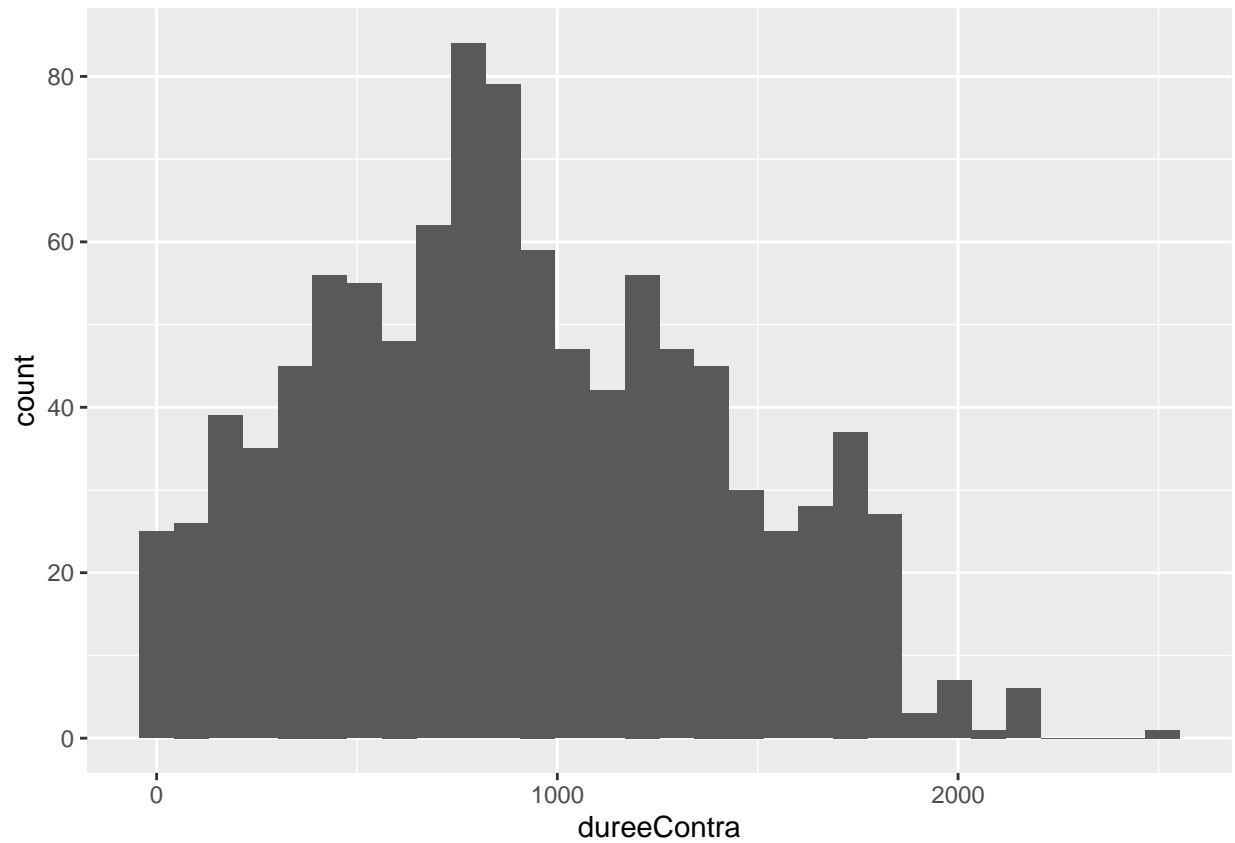
```
##          IDPOL    DATEEFF      IDCLI      IDPART      IDCHI1      IDCHI2 IDCHI3
## 1: 2C7EAEF380 2011-02-10 7A61ADDB61      <NA>      <NA>      <NA> <NA>
## 2: F1125A72F9 2011-07-10 4E5D3AA63C 32D5615B20 2EF179B6C7 B09D69E316 <NA>
## 3: 0F5BF8EA40 2011-08-10 283B50C751      <NA>      <NA>      <NA> <NA>
## 4: 73FB728155 2011-08-10 2DF3A65E55      <NA>      <NA>      <NA> <NA>
## 5: 21BFCE62CB 2012-01-01 9C968CCD1F      <NA>      <NA>      <NA> <NA>
## 6: 0EF51BD7B0 2011-12-10 51305185AA      <NA>      <NA>      <NA> <NA>
##      IDCHI4 IDCHI5 IDCHI6 FORMU      DATECAN      REACAN AMTPREM
## 1:  <NA>  <NA>      NA  FOR1 2017-12-31 Non-payment of the premium 1455
## 2:  <NA>  <NA>      NA  FOR4 2016-12-31 Non-payment of the premium 4242
## 3:  <NA>  <NA>      NA  FOR2 2016-12-31 Non-payment of the premium 1977
## 4:  <NA>  <NA>      NA  FOR3 2016-12-31 Non-payment of the premium 3628
## 5:  <NA>  <NA>      NA  FOR4 2016-12-31 Non-payment of the premium 4738
## 6:  <NA>  <NA>      NA  FOR4 2016-12-31 Non-payment of the premium 4320
##      RANKAGG      COMPAR  MARCLI SEXCLI      DTNCLI  PROFCLI      CP SEXCJ
## 1:      1 Comparateur1 Single      M 1980-12-16 Employee 29200 <NA>
## 2:      4 Comparateur1 Married      M 1968-12-22 Employee 1700      F
## 3:      1 Comparateur1 Single      M 1977-06-08 Employee 45260 <NA>
## 4:      2 Comparateur1 Single      M 1960-09-05 Employee 74130 <NA>
## 5:      2 Comparateur1 Single      F 1951-02-19 Retired 14600 <NA>
## 6:      6 Comparateur1 Single      F 1962-07-12 Employee 66600 <NA>
##      DTNCJ  PROF CJ nombre_enfants resilie dureeContra
## 1:  <NA>  <NA>      0      Oui 2516 days
## 2: 1978-11-15 Employee      2      Oui 2001 days
## 3:  <NA>  <NA>      0      Oui 1970 days
## 4:  <NA>  <NA>      0      Oui 1970 days
## 5:  <NA>  <NA>      0      Oui 1826 days
## 6:  <NA>  <NA>      0      Oui 1848 days
```

Vérification

```
ggplot()+geom_histogram(data=contra_resil_clean,aes(x=dureeContra))
```

```
## Don't know how to automatically pick scale for object of type difftime. Defaulting to continuous.
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

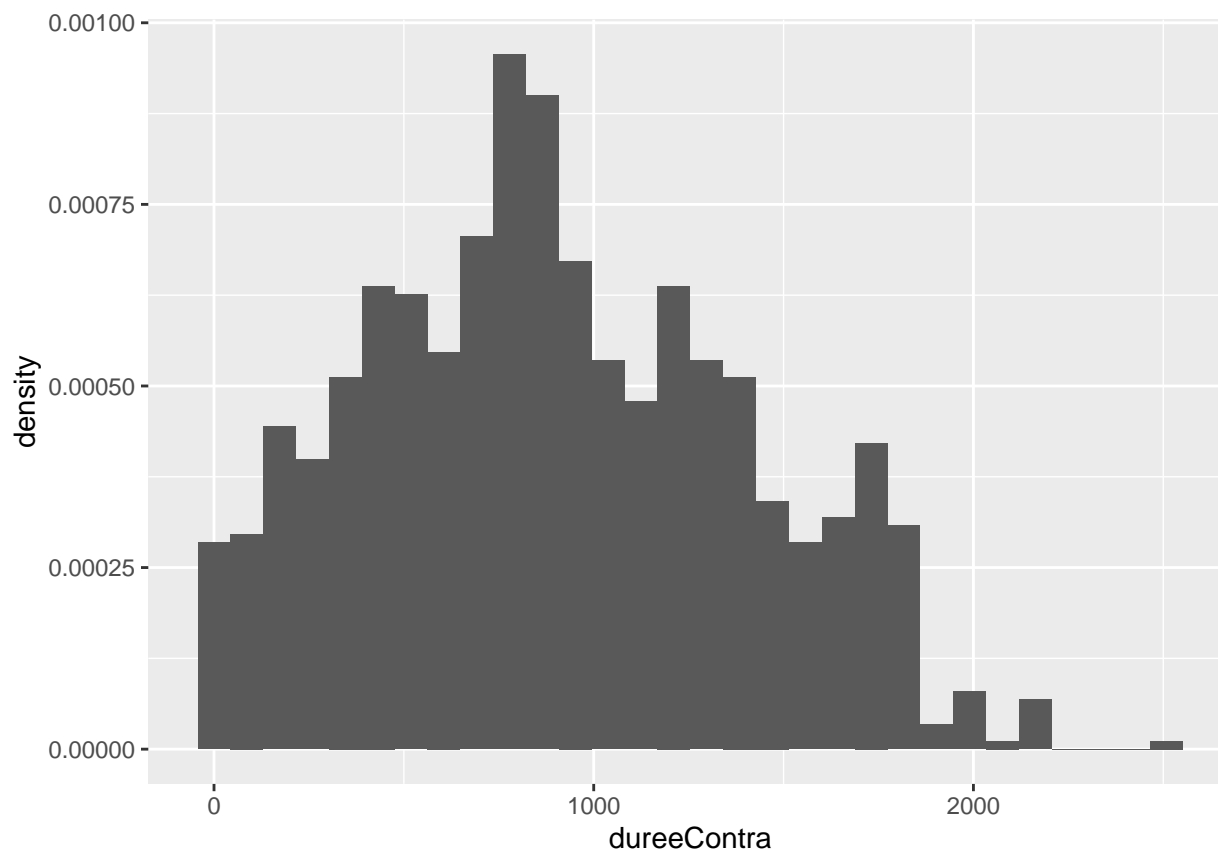


```
ggplot()+geom_histogram(data=contra_resil_clean,aes(x=dureeContra,y=..density..))+geom_density()
```

```
## Don't know how to automatically pick scale for object of type difftime. Defaulting to continuous.
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```





- Analyse de la durée de la contrats par formule:

```
contra_resil_clean[,.(moyenne_duree=mean(dureeContra),moyenne_primes=mean(AMTPREM)),by=FORMU]
```

```
##      FORMU  moyenne_duree moyenne_primes
## 1:  FOR1 1058.8476 days      2417.229
## 2:  FOR4  869.7993 days      2392.611
## 3:  FOR2 1102.8367 days      2319.837
## 4:  FOR3  885.9118 days      2521.294
```

#### Interprétation:

- Les clients qui ont souscrit aux formules 1 et 2 sont en général plus fidèles et ont des durées de contrats plus longues.
- Cela s'explique vu la moyenne des primes. Les formules 1 et 2 sont les formules les moins chères et donc les clients sont en général plus fidèles.

```
contra_resil_clean[,.(moyenne_duree=mean(dureeContra),moyenne_primes=mean(AMTPREM)),by=.(COMPAR)]
```

```
##          COMPAR  moyenne_duree moyenne_primes
## 1: Comparateur1 1064.6655 days      2317.109
## 2: Comparateur2  891.1818 days      2398.665
## 3: Comparateur3  924.5068 days      2323.438
## 4: Comparateur4  870.5975 days      2296.201
## 5: Comparateur5  784.8833 days      2422.417
## 6: Comparateur6  779.5037 days      2545.853
```

#### Interprétations:

- Le comparateur 1, bien que pas très convoité par les clients, est le comparateur où les contrats sont plus longs, cela s'explique sûrement vu les prix des primes de chaque comparateur.
- les clients souscrivent le plus via les comparateurs 6 et 5, bien qu'offrant des primes relativement élevées.

## 6. fusion des tables contrats et prestations

Etude des montants des dépenses en fonction de plusieurs variables

```
contra_prest=merge(contra_comp, prest, by = "IDPOL")
```

```
head(contra_prest)
```

```
##          IDPOL    DATEEFF      IDCLI      IDPART      IDCHI1      IDCHI2 IDCHI3
## 1: 000B46F609 2018-06-01 C1B23F6425 52BA84B9C0 6C1EE86F8C D04BB97D0B <NA>
## 2: 000B46F609 2018-06-01 C1B23F6425 52BA84B9C0 6C1EE86F8C D04BB97D0B <NA>
## 3: 000B46F609 2018-06-01 C1B23F6425 52BA84B9C0 6C1EE86F8C D04BB97D0B <NA>
## 4: 000B46F609 2018-06-01 C1B23F6425 52BA84B9C0 6C1EE86F8C D04BB97D0B <NA>
## 5: 000B46F609 2018-06-01 C1B23F6425 52BA84B9C0 6C1EE86F8C D04BB97D0B <NA>
## 6: 000B46F609 2018-06-01 C1B23F6425 52BA84B9C0 6C1EE86F8C D04BB97D0B <NA>
##          IDCHI4 IDCHI5 IDCHI6 FORMU DATECAN REACAN AMTPREM RANKAGG      COMPAR
## 1: <NA> <NA>      NA FOR4 <NA> <NA>      296      NA Comparateur6
## 2: <NA> <NA>      NA FOR4 <NA> <NA>      296      NA Comparateur6
## 3: <NA> <NA>      NA FOR4 <NA> <NA>      296      NA Comparateur6
## 4: <NA> <NA>      NA FOR4 <NA> <NA>      296      NA Comparateur6
## 5: <NA> <NA>      NA FOR4 <NA> <NA>      296      NA Comparateur6
## 6: <NA> <NA>      NA FOR4 <NA> <NA>      296      NA Comparateur6
##          MARCLI SEXCLI      DTNCLI      PROFCLI      CP SEXCJ      DTNCJ
## 1: Married      M 1969-06-07 Without profession 61270      F 1970-06-17
## 2: Married      M 1969-06-07 Without profession 61270      F 1970-06-17
## 3: Married      M 1969-06-07 Without profession 61270      F 1970-06-17
## 4: Married      M 1969-06-07 Without profession 61270      F 1970-06-17
## 5: Married      M 1969-06-07 Without profession 61270      F 1970-06-17
## 6: Married      M 1969-06-07 Without profession 61270      F 1970-06-17
##          PROFCJ nombre_enfants resilie      IDBENE      Categories
## 1: Without profession      2      Non C1B23F6425 Soins ambulatoires
## 2: Without profession      2      Non C1B23F6425 Soins ambulatoires
## 3: Without profession      2      Non C1B23F6425 Soins ambulatoires
## 4: Without profession      2      Non C1B23F6425 Soins ambulatoires
## 5: Without profession      2      Non C1B23F6425      Pharmacie
## 6: Without profession      2      Non C1B23F6425      Pharmacie
##          Subcat      Name      Amt
## 1: MEDECINS CONSULTATION SPECIALISTE MEDECINE GENERALE 25.00
## 2: MEDECINS CONSULTATION SPECIALISTE MEDECINE GENERALE 25.00
## 3: MEDECINS      CONSULTATION SPECIALISTE 50.00
## 4: MEDECINS CONSULTATION SPECIALISTE MEDECINE GENERALE 25.00
## 5: PHARMACIE      PHARMACIE VIGNETTE BLANCHE (70%) 17.84
## 6: PHARMACIE      HONO DISP 7 1.02
```

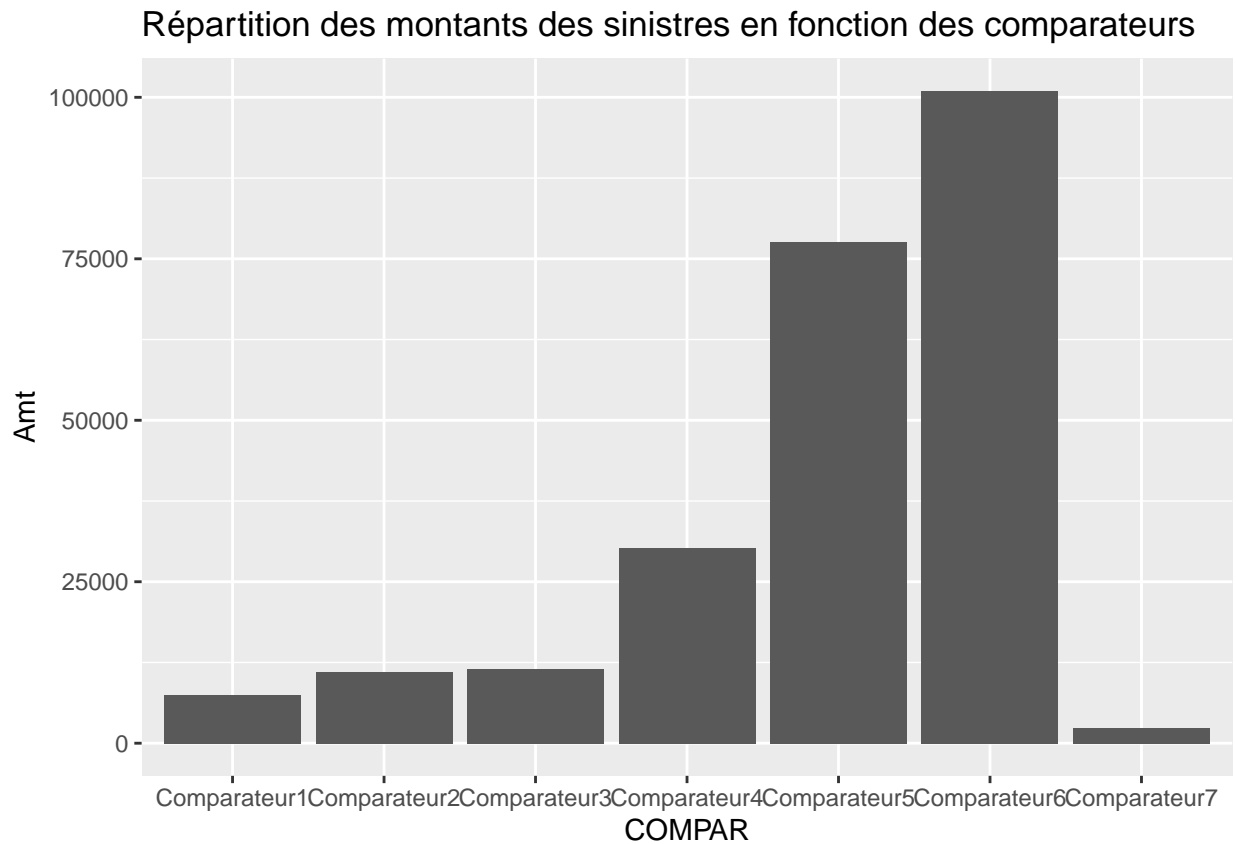
```
dep_compa=contra_prest[,.(moyenne_dep=mean(Amt)),by=COMPAR]
```

```
print(dep_compa)
```

```
##          COMPAR moyenne_dep
## 1: Comparateur6 17.71492
## 2: Comparateur5 18.48737
```

```
## 3: Comparateur4    21.05894
## 4: Comparateur1    13.61505
## 5: Comparateur3    42.21789
## 6: Comparateur2    26.55051
## 7: Comparateur7    25.99386
```

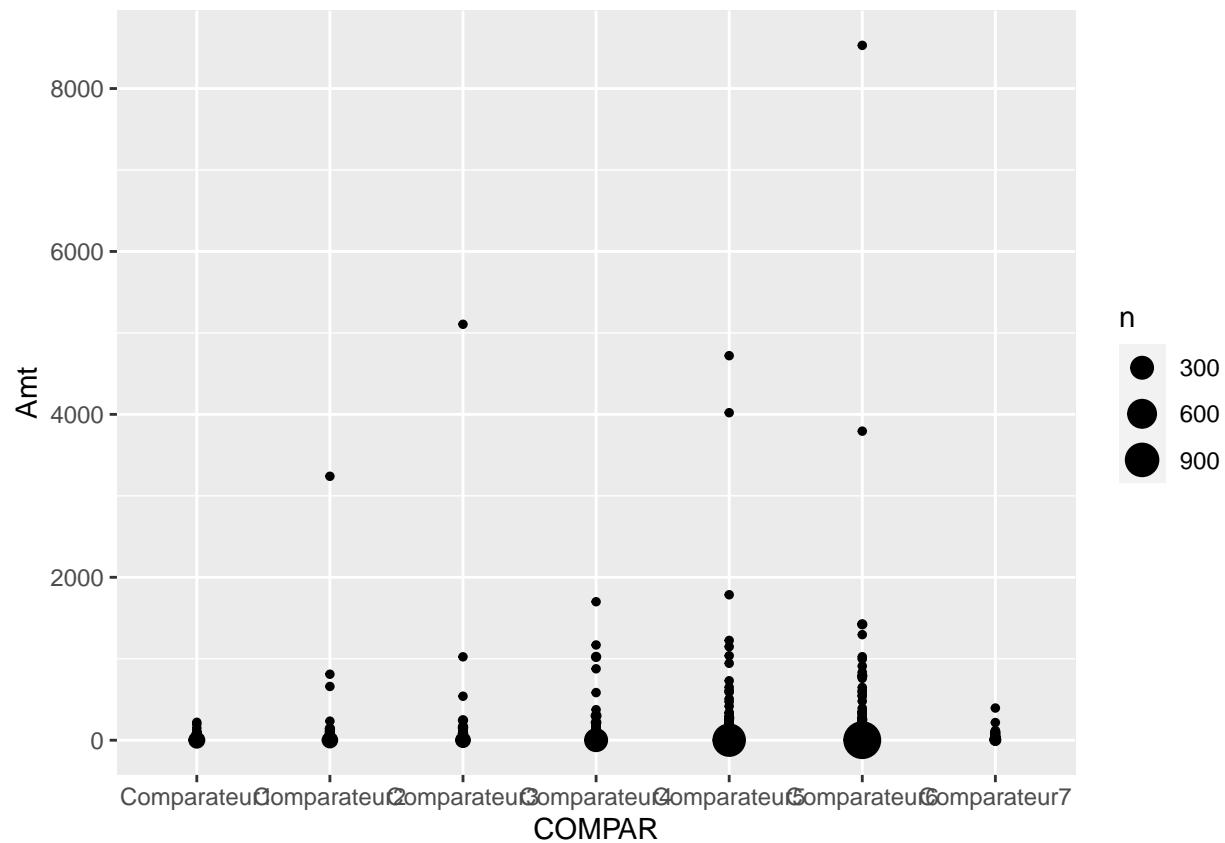
```
ggplot(contra_prest, aes(x=COMPAR,y=Amt))+geom_bar(stat="identity")+ggtitle("Répartition des montants d
```



### Interprétation

- Les clients ayant choisi les comparateurs 4,5,6 sont ceux ayant les montants des sinistres les plus élevés.
- on peut aussi remarquer que le montant des dépenses évolue en fonction du comparateur à l'exception du comparateur 7.

```
ggplot(contra_prest, aes(x=COMPAR,y=Amt))+geom_count()
```

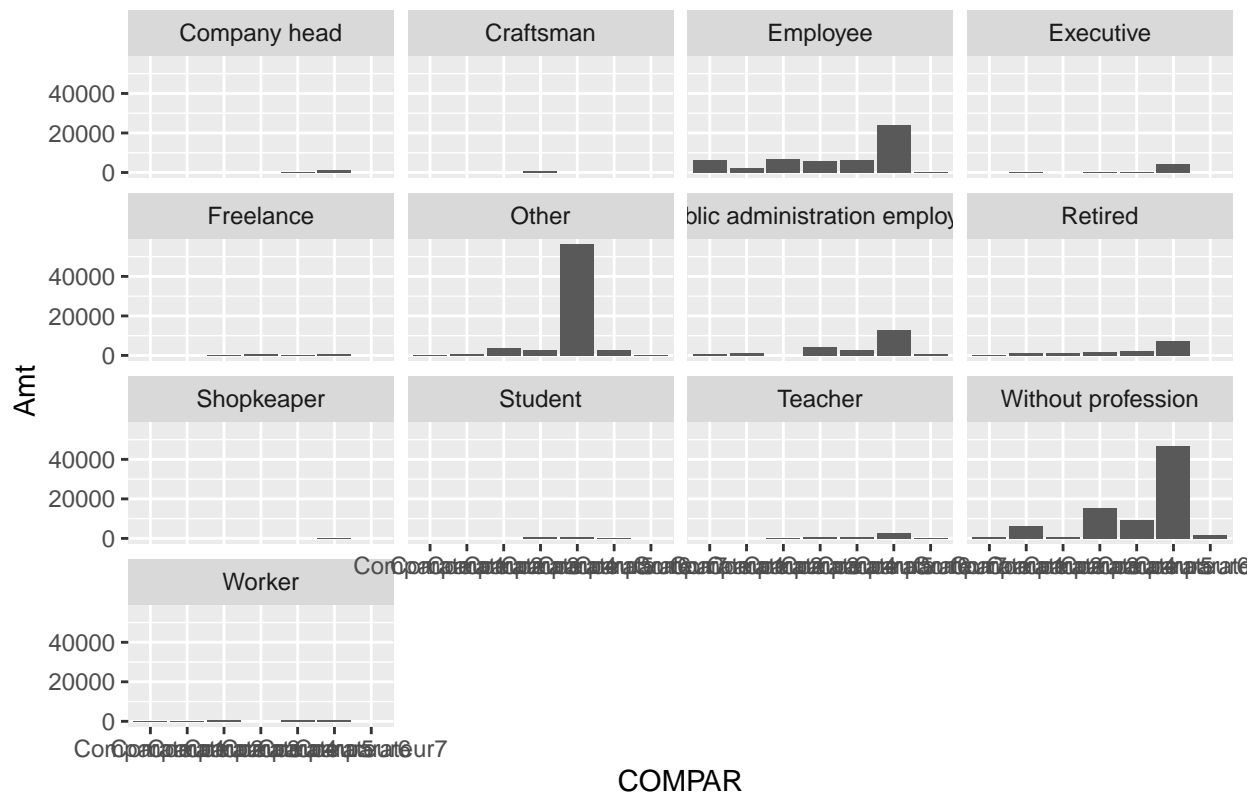


**\*\* Interprétation\*\***

- Ce graphique confirme bien les observations relevées concernant les montants des dépenses par comparateur.

```
ggplot(contra_prest, aes(x=COMPAR,y=Amt))+geom_bar(stat="identity")+facet_wrap(~PROFCLI)+ggtitle("Répar
```

## Répartition des dépenses par comparateurs et par profession



### Interprétation

- on remarque qu'importe la profession , le comparateur 6 est celui qui génère le plus de dépenses à l'exception des personnes ayant "autre" comme profession .
- les employés , sans profession et autres sont ceux qui génèrent le plus de dépenses peu importe le comparateur.

```
tab_depen=contra_prest[,.(moyenne_dep=mean(Amt)),by=MARCLI]
```

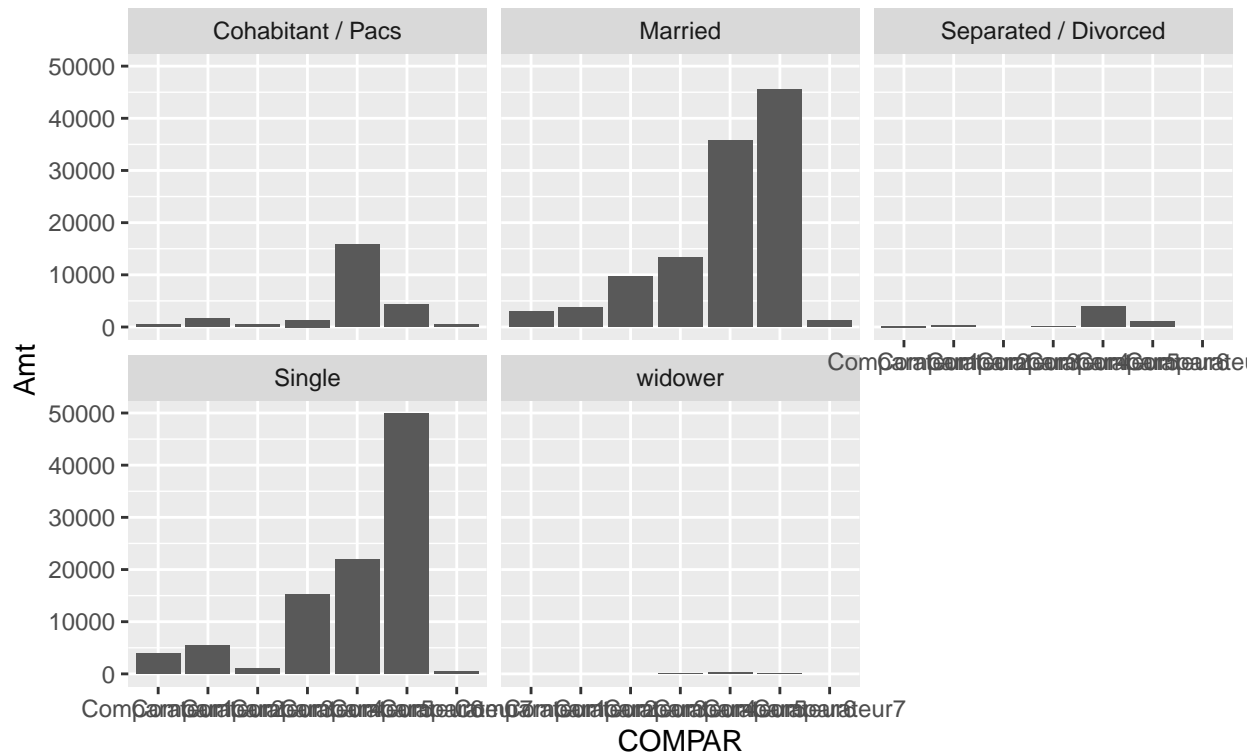
```
print(tab_depen)
```

```
##           MARCLI moyenne_dep
## 1:           Married   18.593123
## 2:           Single   20.276958
## 3: Cohabitant / Pacs   18.765869
## 4: Separated / Divorced 13.352569
## 5:           widower    6.387708
```

```
ggplot(contra_prest, aes(x=COMPAR,y=Amt))+
```

```
geom_bar(stat="identity")+facet_wrap(~MARCLI)+ggtitle("Analyse des sinistres en fonction des comparateur")
```

## Analyse des sinistres en fonction des comparateurs et du statut marital du client



### Interprétation

- Les célibataires et les mariés sont les catégories qui génèrent le plus de dépenses . Celles-ci sont plus élevées pour le comparateur 6 par ailleurs.
- Les couples de la catégorie “Cohabitant/pacs” et “séparés/divorcé” ont des dépenses plus élevées pour ceux ayant choisi le comparateur 5.