

Projet statistique :

Charger les données :

Installer les packages nécessaires :

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

Charger les données :

```
# xlsx files
library(xlsx)

## Warning: package 'xlsx' was built under R version 3.4.2
## Loading required package: rJava
## Warning: package 'rJava' was built under R version 3.4.2
## Loading required package: xlsxjars
## Warning: package 'xlsxjars' was built under R version 3.4.2
my_data <- read.xlsx("dataset.xlsx",sheetIndex = 1,as.data.frame = T,encoding = 'UTF-8')
```

nommer les colonnes :

```
my_col_names=c('date','a_suivi','performer_travail_projet_1','motivation_certificat_1','trouver_emploi_1')
colnames(my_data)<-my_col_names
```

Choisir lignes intéressants :

```
## j'ai nettoyé les lignes à la main donc on prend tout les lignes
my_data2<-my_data[c(1:nrow(my_data)),]
```

Eliminer les colonnes : mail, date

```
my_data2$date<-NULL
my_data2$mail<-NULL
my_data2$commentaire<-NULL
```

créer les dataframes nécessaires:

librairie pour filtrer les lignes :

```
library("dplyr")
```

```
## Warning: package 'dplyr' was built under R version 3.4.2
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

str(my_data2)

## 'data.frame':   134 obs. of  52 variables:
##  $ a_suivi          : Factor w/ 3 levels "Jamais","Oui, je l'ai suivi et j'ai réussi",...: 2
##  $ performer_travail_projet_1: Factor w/ 4 levels "D'accord","Désaccord",...: NA NA NA NA NA NA NA NA NA
##  $ motivation_certificat_1   : Factor w/ 4 levels "D'accord","Désaccord",...: NA NA NA NA NA NA NA NA NA
##  $ trouver_emploi_1         : Factor w/ 4 levels "D'accord","Désaccord",...: NA NA NA NA NA NA NA NA NA
##  $ mooc_inclu_etude_1       : Factor w/ 2 levels "Non","Oui": NA NA NA NA NA NA NA NA NA ...
##  $ edt_chargee_1           : Factor w/ 2 levels "Non","Oui": NA NA NA NA NA NA NA NA NA ...
##  $ plateformes_connues       : Factor w/ 9 levels "Coursera","Coursera, Edx",...: NA NA NA NA NA NA NA NA NA
##  $ aide_financiere_1        : Factor w/ 2 levels "Non","Oui": NA NA NA NA NA NA NA NA NA ...
##  $ prix_1                  : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ outils_exercices_1       : Factor w/ 1 level "Oui": NA NA NA NA NA NA NA NA NA ...
##  $ sciences_donnees_1       : Factor w/ 3 levels "Inintéressant",...: NA NA NA NA NA NA NA NA NA NA
##  $ informatique_1           : Factor w/ 3 levels "Inintéressant",...: NA NA NA NA NA NA NA NA NA NA
##  $ sciences_physiques_1     : Factor w/ 3 levels "Inintéressant",...: NA NA NA NA NA NA NA NA NA NA
##  $ business_1              : Factor w/ 3 levels "Inintéressant",...: NA NA NA NA NA NA NA NA NA NA
##  $ arts_1                  : Factor w/ 3 levels "Inintéressant",...: NA NA NA NA NA NA NA NA NA NA
##  $ langues_1                : Factor w/ 2 levels "Interessant",...: NA NA NA NA NA NA NA NA NA NA ..
##  $ sciences_sociales_1      : Factor w/ 3 levels "Inintéressant",...: NA NA NA NA NA NA NA NA NA NA
##  $ performer_travail_projet   : Factor w/ 5 levels "D'accord","Désaccord",...: 4 4 1 1 5 3 4 1 4 1 ...
##  $ motivation_certificat     : Factor w/ 5 levels "D'accord","Désaccord",...: 1 4 1 2 5 1 2 1 2 4 ...
##  $ trouver_emploi           : Factor w/ 5 levels "D'accord","Désaccord",...: 3 4 1 5 5 2 3 1 2 2 ...
##  $ nbr_certificat           : num  1 2 1 1 1 1 1 1 3 1 ...
##  $ satisfait_gestion_temps    : Factor w/ 5 levels "D'accord","Désaccord",...: 1 4 1 3 4 4 2 1 4 3 ...
##  $ objectifs_atteints        : Factor w/ 5 levels "D'accord","Désaccord",...: 4 4 1 1 3 1 5 1 4 1 ...
##  $ satisfaction              : num  7 10 9 6 1 7 3 6 7 6 ...
##  $ vacances_etude            : Factor w/ 3 levels "Les deux","Les vacances",...: 1 1 1 3 3 3 1 3 1 1
##  $ temps_alloue_cours        : Factor w/ 18 levels "0","0.01","0.1",...: 14 9 13 5 9 11 9 7 13 5 ...
##  $ mooc_inclu_etude          : Factor w/ 2 levels "Non","Oui": 1 1 2 2 2 2 2 2 1 ...
##  $ edt_chargee               : Factor w/ 2 levels "Non","Oui": 1 1 2 2 2 2 2 1 1 2 ...
##  $ plateformes_utilisees      : Factor w/ 30 levels "Canvas, OpenClassrooms, FUN",...: 14 2 21 23 23 2
##  $ videos_cours              : Factor w/ 5 levels "Bien","Mauvais",...: 4 4 1 3 1 3 1 4 4 1 ...
##  $ transcription             : Factor w/ 5 levels "Bien","Mauvais",...: 1 4 1 1 1 3 3 4 1 4 ...
##  $ diapositives_cours        : Factor w/ 5 levels "Bien","Mauvais",...: 1 4 1 1 3 1 4 1 4 1 ...
##  $ forum_discussion          : Factor w/ 5 levels "Bien","Mauvais",...: 1 4 1 4 3 3 5 1 4 1 ...
##  $ examens_exercices         : Factor w/ 5 levels "Bien","Mauvais",...: 4 4 1 3 2 1 1 3 1 1 ...
##  $ accompagnement_pedagogique: Factor w/ 5 levels "Bien","Mauvais",...: 1 4 1 3 1 2 1 1 1 3 ...
##  $ aides_financiere          : Factor w/ 2 levels "Non","Oui": 2 2 2 1 1 1 2 1 1 1 ...
##  $ prix                      : num  NA 10 20 0 0 0 0 100 50 30 ...
##  $ outils_exercice           : Factor w/ 2 levels "Non","Oui": 2 2 2 1 2 2 2 2 1 2 ...
##  $ sciences_donnees          : Factor w/ 3 levels "Inintéressant",...: 2 2 2 2 2 3 2 2 2 2 ...
##  $ informatique              : Factor w/ 3 levels "Inintéressant",...: 2 2 3 2 3 2 2 2 2 2 ...
##  $ sciences_physiques        : Factor w/ 3 levels "Inintéressant",...: 2 3 2 2 2 2 3 2 3 3 ...
##  $ business                  : Factor w/ 3 levels "Inintéressant",...: 2 3 3 3 1 2 3 1 1 1 ...
```

```
## $ arts : Factor w/ 3 levels "Inintéressant",...: 1 3 2 1 1 3 2 1 3 3 ...
## $ langues : Factor w/ 3 levels "Inintéressant",...: 1 3 2 1 1 2 1 1 3 3 ...
## $ sciences_sociales : Factor w/ 3 levels "Inintéressant",...: 1 3 3 1 1 2 1 1 1 1 ...
## $ sexe : Factor w/ 2 levels "Femme","Homme": 2 2 2 2 2 1 2 2 2 2 ...
## $ pays : Factor w/ 13 levels "Autre","Burkinafaso",...: 8 8 5 5 5 5 8 5 5 5 ..
## $ age : num 22 22 22 22 20 22 22 22 23 23 ...
## $ est_etudiant : Factor w/ 2 levels "Non","Oui": 2 2 2 2 2 2 2 2 2 2 ...
## $ niveau_actuel : Factor w/ 5 levels "BAC +2","BAC +3",...: 4 4 2 4 2 3 4 4 3 4 ...
## $ formation : Factor w/ 6 levels "FIG","FIL","FIP",...: 1 1 1 1 1 1 1 1 1 1 ...
## $ background : Factor w/ 4 levels "Autre école",...: 3 3 4 3 3 3 3 3 3 3 ...
```

```
#change from factor to numeric the feature : temps_alloue_cours
my_data2$temps_alloue_cours<-as.numeric(my_data2$temps_alloue_cours)
```

jamais suivi :

```
data_jamais<-filter(my_data2,a_suivi=="Jamais")
# supprimer les colonnes avec que de Nans :
data_jamais<-data_jamais[ , ! apply( data_jamais , 2 , function(x) all(is.na(x)) ) ]
# Write CSV in R
write.csv(data_jamais, file = "data_jamais.csv")
```

```
str(data_jamais)
```

```
## 'data.frame': 19 obs. of 24 variables:
## $ a_suivi : Factor w/ 3 levels "Jamais","Oui, je l'ai suivi et j'ai réussi",...: 1
## $ performer_travail_projet_1: Factor w/ 4 levels "D'accord","Désaccord",...: 3 1 3 3 3 4 1 4 4 1 ..
## $ motivation_certificat_1 : Factor w/ 4 levels "D'accord","Désaccord",...: 1 1 3 1 1 1 4 1 4 1 ...
## $ trouver_emploi_1 : Factor w/ 4 levels "D'accord","Désaccord",...: 4 2 3 2 2 3 1 3 4 3 ...
## $ mooc_inclu_etude_1 : Factor w/ 2 levels "Non","Oui": 2 2 2 2 2 1 2 2 1 ...
## $ edt_chargée_1 : Factor w/ 2 levels "Non","Oui": 1 1 1 2 2 1 2 2 1 2 ...
## $ plateformes_connues : Factor w/ 9 levels "Coursera","Coursera, Edx",...: 8 8 1 4 7 5 3 6 8 2
## $ aide_financiere_1 : Factor w/ 2 levels "Non","Oui": 1 1 1 2 1 1 2 1 2 2 ...
## $ prix_1 : num NA 40 NA 50 0 NA 20 20 5 5 ...
## $ outils_exercices_1 : Factor w/ 1 level "Oui": 1 1 1 1 1 1 1 1 1 1 ...
## $ sciences_donnees_1 : Factor w/ 3 levels "Inintéressant",...: 2 3 3 2 2 2 2 3 2 2 ...
## $ informatique_1 : Factor w/ 3 levels "Inintéressant",...: 2 2 2 3 2 2 2 3 2 2 ...
## $ sciences_physiques_1 : Factor w/ 3 levels "Inintéressant",...: 3 2 2 3 2 2 2 2 2 1 ...
## $ business_1 : Factor w/ 3 levels "Inintéressant",...: 2 2 2 2 2 2 1 3 2 3 ...
## $ arts_1 : Factor w/ 3 levels "Inintéressant",...: 3 3 3 1 2 3 2 1 2 2 ...
## $ langues_1 : Factor w/ 2 levels "Interessant",...: 1 1 1 1 1 1 1 1 2 1 1 ...
## $ sciences_sociales_1 : Factor w/ 3 levels "Inintéressant",...: 3 2 3 1 2 2 3 3 2 2 ...
## $ sexe : Factor w/ 2 levels "Femme","Homme": 2 2 2 2 1 2 2 2 2 2 ...
## $ pays : Factor w/ 13 levels "Autre","Burkinafaso",...: 4 5 11 1 5 5 5 5 4 8 .
## $ age : num 20 23 32 28 21 20 22 20 35 21 ...
## $ est_etudiant : Factor w/ 2 levels "Non","Oui": 2 2 2 2 2 2 2 2 2 2 ...
## $ niveau_actuel : Factor w/ 5 levels "BAC +2","BAC +3",...: 4 4 5 4 2 1 3 2 4 3 ...
## $ formation : Factor w/ 6 levels "FIG","FIL","FIP",...: 1 1 6 5 1 3 1 3 6 1 ...
## $ background : Factor w/ 4 levels "Autre école",...: 3 3 4 4 4 2 4 2 4 3 ...
```

suivi terminé et pas terminé :

```
data_suivi<-filter(my_data2,a_suivi %in% c("Oui, je l'ai suivi mais je ne l'ai pas terminé", "Oui, je l'ai suivi et j'ai réussi"))
data_suivi<-data_suivi[ , ! apply( data_suivi , 2 , function(x) all(is.na(x)) ) ]
```

```
write.csv(data_suivi, file = "data_suivi.csv")
```

suivi mais pas terminé :

```
data_non_termine<-filter(my_data2,a_suivi %in% c("Oui, je l'ai suivi mais je ne l'ai pas terminé"))
data_non_termine<-data_non_termine[ , ! apply( data_non_termine , 2 , function(x) all(is.na(x)) ) ]
write.csv(data_non_termine, file = "data_non_termine.csv")
```

suivi et terminé :

```
data_termine<-filter(my_data2,a_suivi %in% c( "Oui, je l'ai suivi et j'ai réussi" ))
data_termine<-data_termine[ , ! apply( data_termine , 2 , function(x) all(is.na(x)) ) ]
write.csv(data_termine, file = "data_termine.csv")
```

Exploration des données : suivi terminé/non terminé

Summary about all features :

```
summary(data_suivi)
```

```
##                               a_suivi
##  Jamais                        : 0
##  Oui, je l'ai suivi et j'ai réussi      :73
##  Oui, je l'ai suivi mais je ne l'ai pas terminé:42
##
##
##
##
##          performer_travail_projet          motivation_certificat
##  D'accord                :51          D'accord                :27
##  Désaccord                : 4          Désaccord                :25
##  Neutre                   :22          Neutre                   :17
##  Tout à fait d'accord      :37          Tout à fait d'accord      :30
##  Tout à fait en désaccord: 1          Tout à fait en désaccord:16
##
##
##
##          trouver_emploi nbr_certificat
##  D'accord                :23      Min.   : 0.000
##  Désaccord                :30      1st Qu.: 0.000
##  Neutre                   :32      Median : 1.000
##  Tout à fait d'accord      : 6      Mean   : 2.157
##  Tout à fait en désaccord:24      3rd Qu.: 1.000
##                                     Max.   :100.000
##
##
##          satisfait_gestion_temps          objectifs_atteints
##  D'accord                :33          D'accord                :44
##  Désaccord                :27          Désaccord                :19
##  Neutre                   :28          Neutre                   :33
##  Tout à fait d'accord      :22          Tout à fait d'accord      :17
##  Tout à fait en désaccord: 5          Tout à fait en désaccord: 2
##
##
```

```

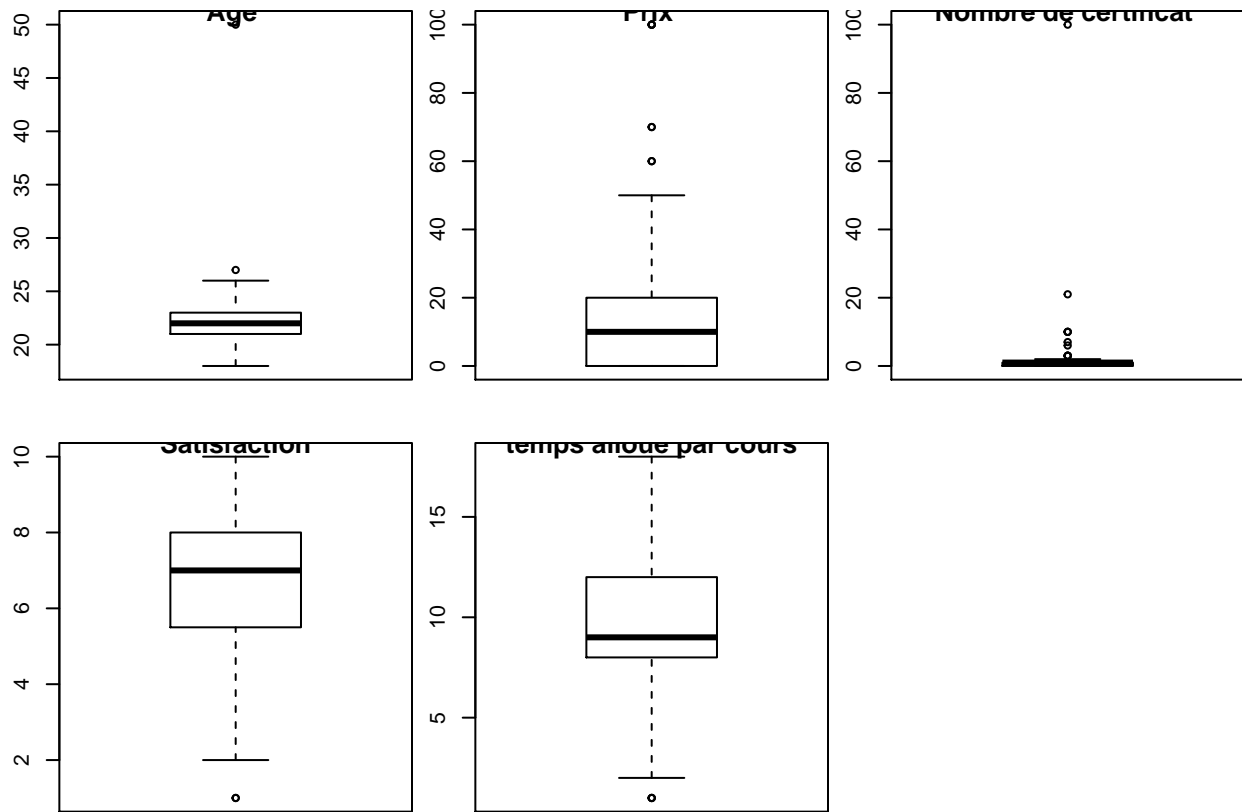
##      satisfaction      vacances_etude temps_alloue_cours mooc_inclu_etude
## Min.   : 1.0    Les deux      :58    Min.   : 1.00    Non:25
## 1st Qu.: 5.5    Les vacances: 9    1st Qu.: 8.00    Oui:90
## Median : 7.0    Vos études   :48    Median : 9.00
## Mean   : 6.6                                Mean   : 9.61
## 3rd Qu.: 8.0                                3rd Qu.:12.00
## Max.   :10.0                                Max.   :18.00
##                                     NA's   :10
##      edt_chargee      plateformes_utilisees      videos_cours
## Non:47    FUN                                :33    Bien      :52
## Oui:68    OpenClassrooms                    :16    Mauvais   : 3
##           OpenClassrooms, FUN                :12    Moyen     :16
##           Coursera                          :11    Très bien :43
##           Coursera, FUN                     : 4    Très mauvais: 1
##           Coursera, OpenClassrooms, FUN: 4
##           (Other)                          :35
##      transcription      diapositives_cours      forum_discussion
## Bien      :52    Bien      :56    Bien      :42
## Mauvais   : 6    Mauvais   : 4    Mauvais   : 9
## Moyen     :21    Moyen     :31    Moyen     :44
## Très bien :35    Très bien :23    Très bien :14
## Très mauvais: 1    Très mauvais: 1    Très mauvais: 6
##
##
##      examens_exercices accompagnement_pedagogique aides_financiere
## Bien      :59    Bien      :41    Non:91
## Mauvais   : 2    Mauvais   :15    Oui:24
## Moyen     :31    Moyen     :36
## Très bien :22    Très bien :15
## Très mauvais: 1    Très mauvais: 8
##
##
##      prix      outils_exercice      sciences_donnees
## Min.   : 0.00    Non:20    Inintéressant : 8
## 1st Qu.: 0.00    Oui:95    Intéressant   :87
## Median :10.00                                Pas très intéressant:20
## Mean   :17.14
## 3rd Qu.:20.00
## Max.   :100.00
## NA's   :19
##      informatique      sciences_physiques
## Inintéressant : 4    Inintéressant :21
## Intéressant   :91    Intéressant   :49
## Pas très intéressant:20    Pas très intéressant:45
##
##
##
##      business      arts
## Inintéressant :21    Inintéressant :37
## Intéressant   :56    Intéressant   :48
## Pas très intéressant:38    Pas très intéressant:30
##
##

```

```
##
##
##          langues          sciences_sociales    sexe
## Inintéressant      :35  Inintéressant      :38    Femme:38
## Intéressant        :50  Intéressant        :38    Homme:77
## Pas très intéressant:30  Pas très intéressant:39
##
##
##
##
##          pays          age          est_etudiant    niveau_actuel
## France      :69  Min.    :18.0  Non: 2      BAC +2      : 7
## Maroc       :23  1st Qu.:21.0  Oui:113   BAC +3      :20
## Tunisie     : 9  Median :22.0           BAC +4      :36
## Autre       : 6  Mean   :22.1           BAC +5      :45
## Côte d'ivoire: 2  3rd Qu.:23.0           BAC +6 ou plus: 4
## Burkinafaso : 1  Max.    :50.0           NA's       : 3
## (Other)     : 5
##          formation          background
## FIG         :97  Autre école: 4
## FIL         : 4  DUT         : 4
## FIP         : 3  Prépa       :95
## Master de recherche: 1  Université :12
## Master of science : 7
## Master spécialisé : 3
##
```

Boxplot pour les variables numériques:

```
par(mfrow=c(2,3),mar=c(2.5,2.5,0,0))
boxplot(data_suivi$age,main="Age ")
boxplot(data_suivi$prix,main="Prix ")
boxplot(data_suivi$nbr_certificat,main="Nombre de certificat ")
boxplot(data_suivi$satisfaction,main="Satisfaction")
boxplot(data_suivi$temps_alloue_cours,main="temps alloué par cours")
```

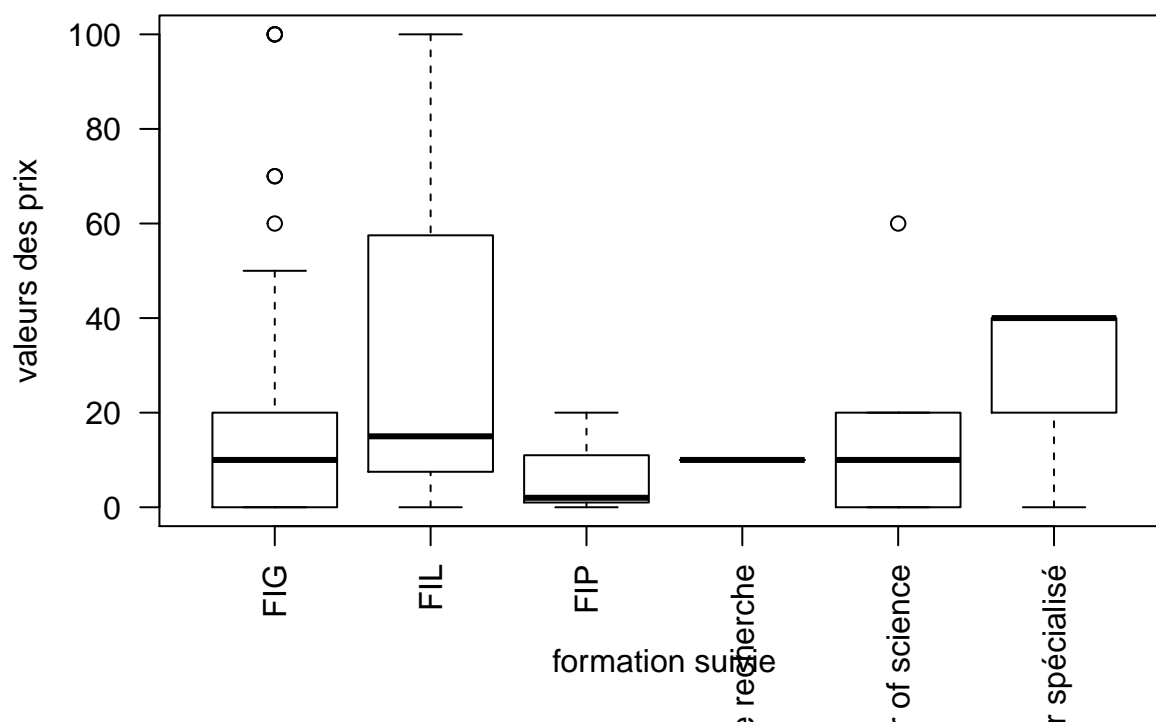


Commentaire : Toutes les variables présentent des outliers surtout prix et nombre de certificat.

exploration des variables qualitatives :

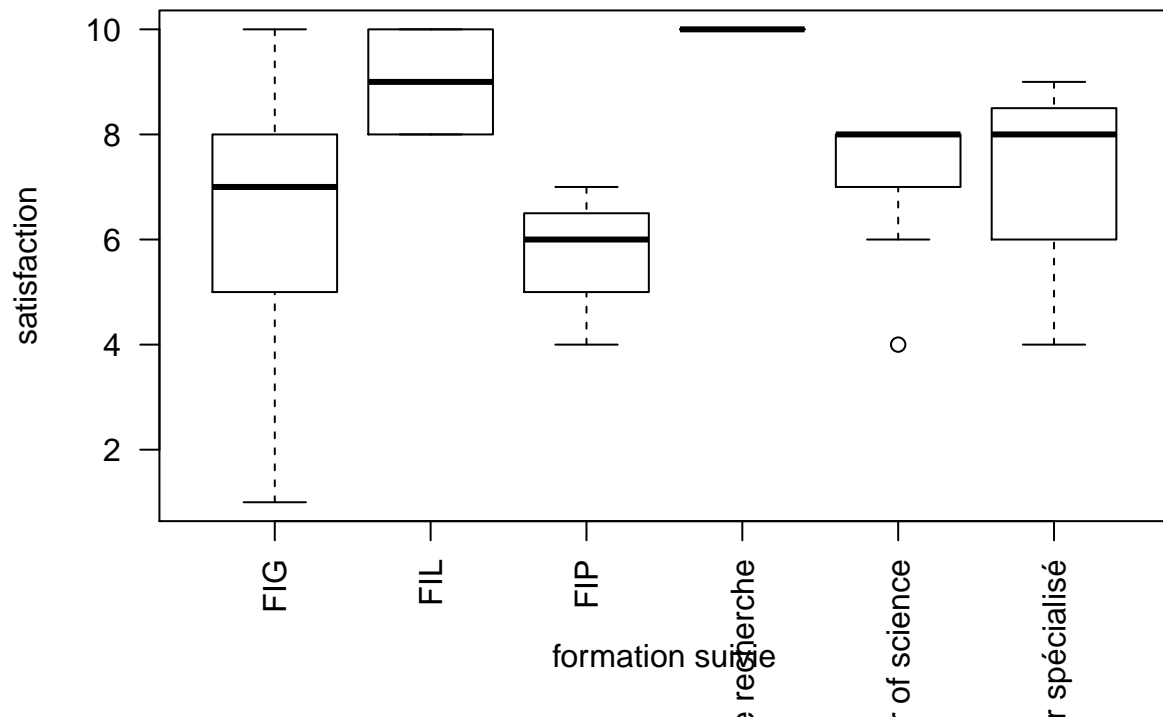
```
boxplot(data_suivi$prix~data_suivi$formation,data=data_suivi, main="Distribution des prix vs formation"
```

Distribution des prix vs formation



```
boxplot(data_suivi$satisfaction~data_suivi$formation,data=data_suivi, main="Distribution de la satisfacti
```

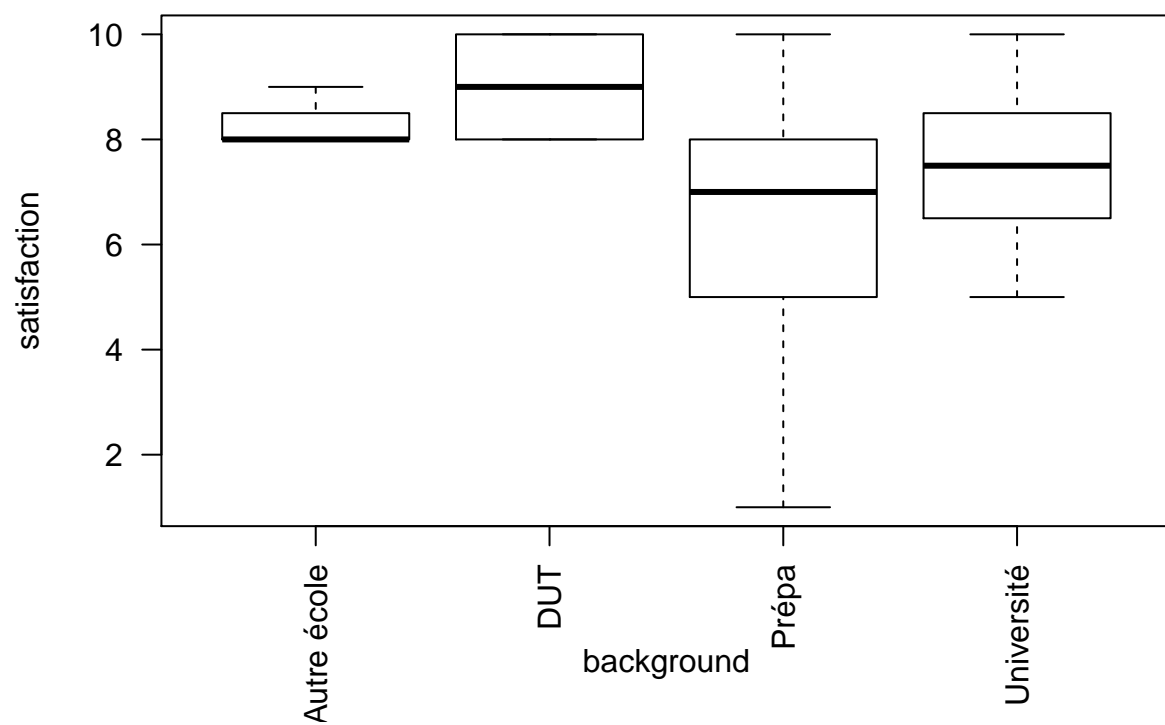

Distribution de la satisfaction vs formation



On peut dire que tout le monde est satisfait de ses moocs.

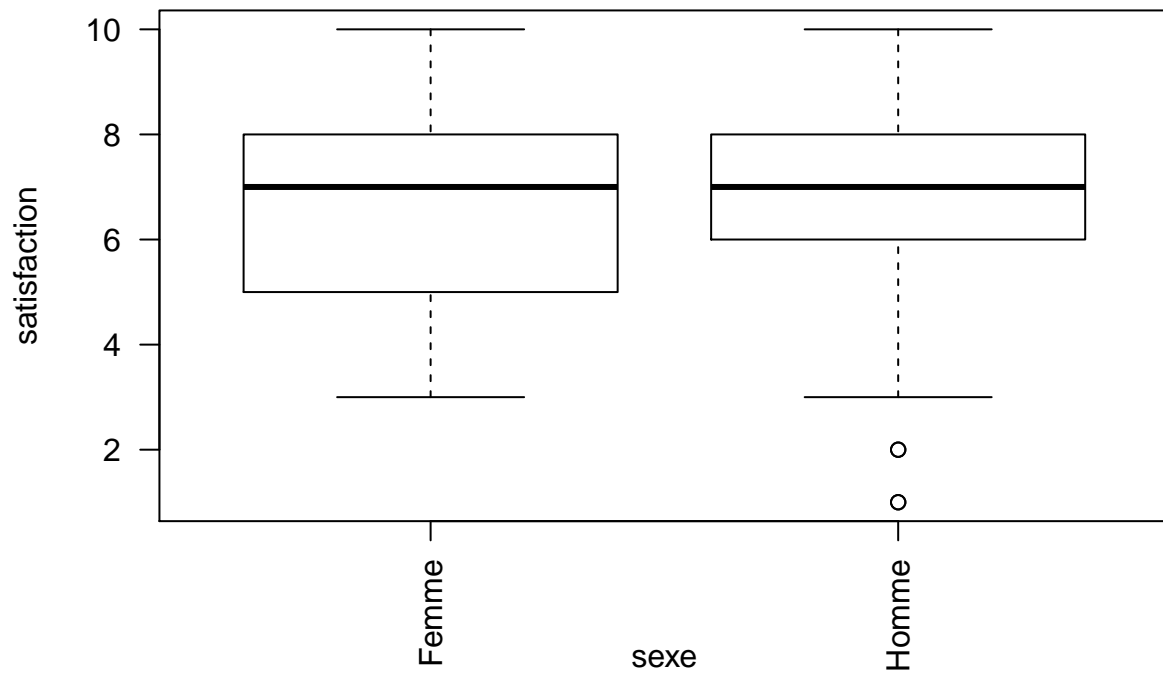
```
boxplot(data_suivi$satisfaction~data_suivi$background,data=data_suivi, main="Distribution de la satisf
```

Distribution de la satisfaction vs background



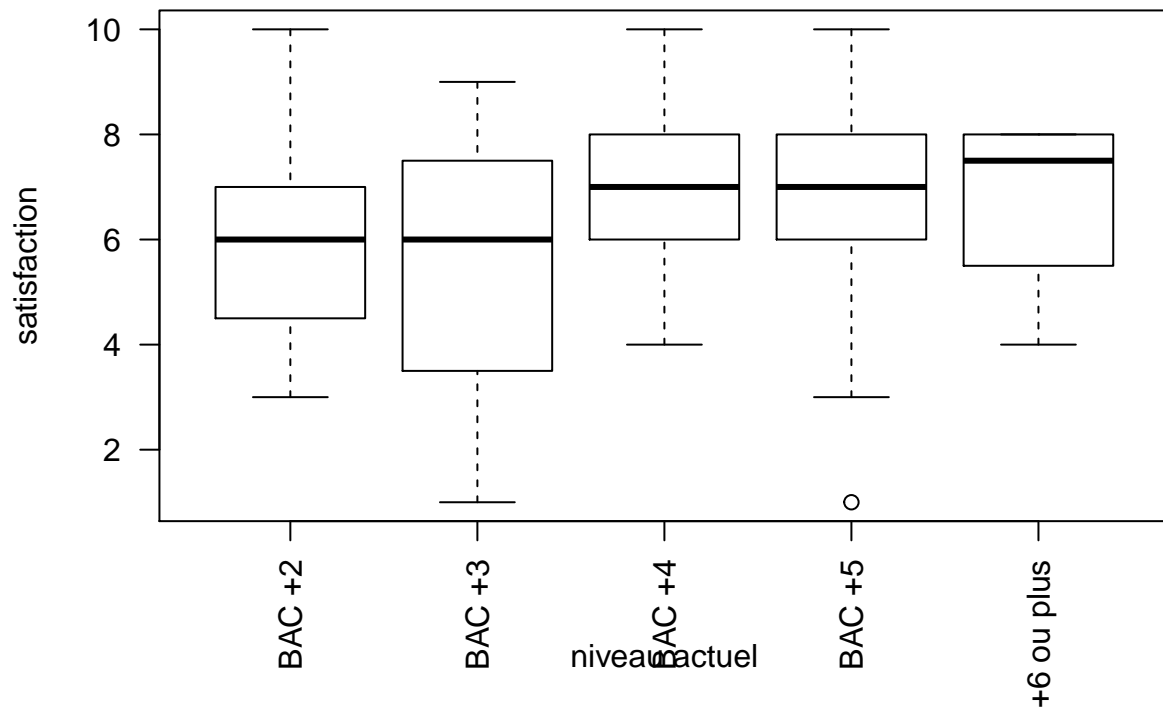
```
boxplot(data_suivi$satisfaction~data_suivi$sexe,data=data_suivi, main="Distribution de la satisfaction v
```

Distribution de la satisfaction vs background



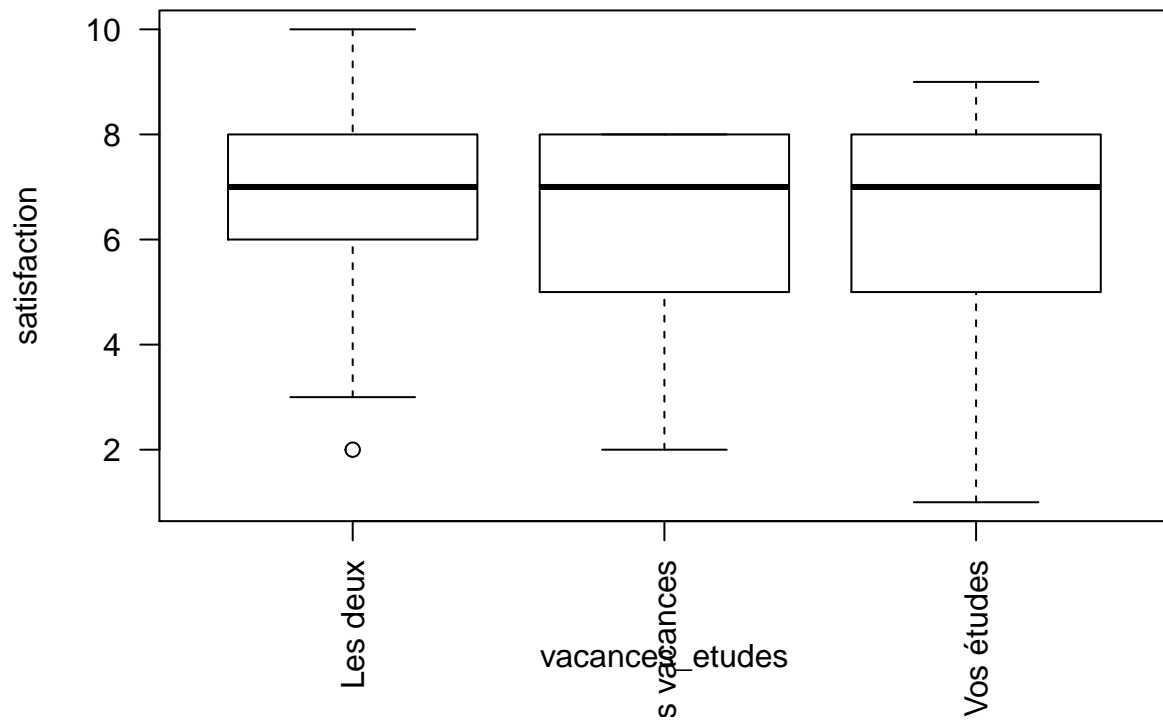
```
boxplot(data_suivi$satisfaction~data_suivi$niveau_actuel,data=data_suivi, main="Distribution de la sati
```

Distribution de la satisfaction vs niveau actuel



```
boxplot(data_suivi$satisfaction~data_suivi$vacances_etude,data=data_suivi, main="Distribution du temps a
```

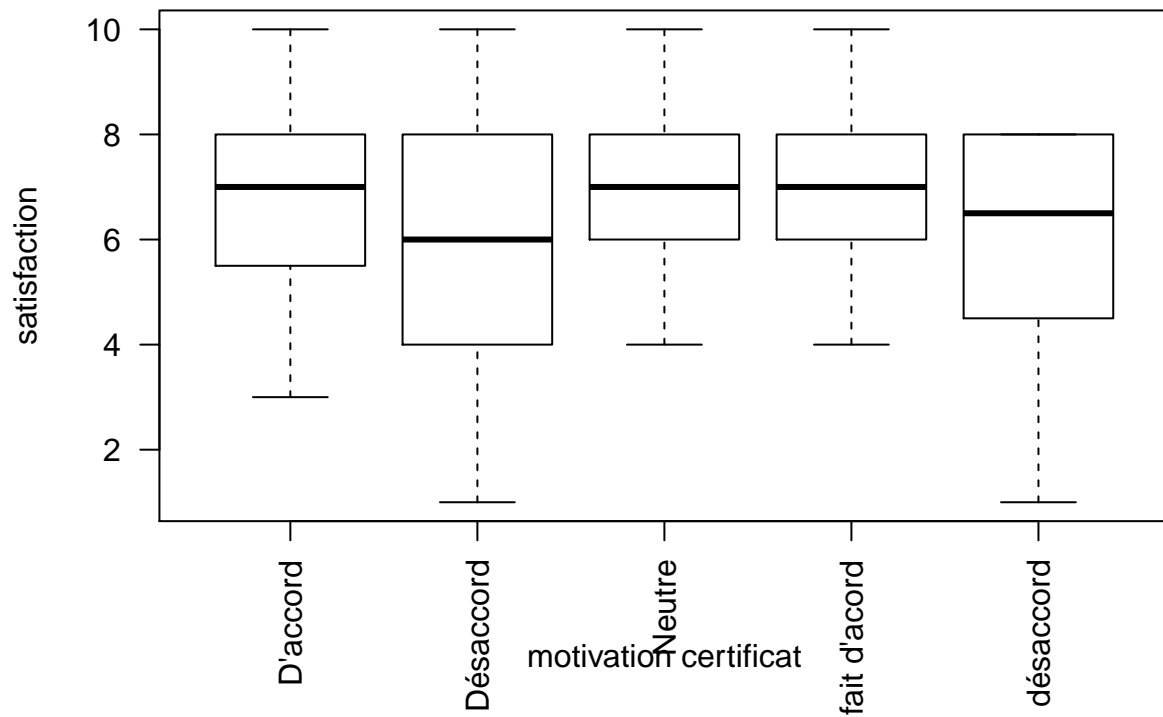
Distribution du temps alloué par cours vs vacances_etudes



```
boxplot(data_suivisatisfaction~data_suivivacances_etude,data=data_suivi, main="Distribution du temps alloué par cours vs vacances_etudes",ylab="satisfaction", xlab="vacances_etudes",las=2)
```

```
boxplot(data_suivi$satisfaction~data_suivi$motivation_certificat,data=data_suivi, main="Distribution du
```

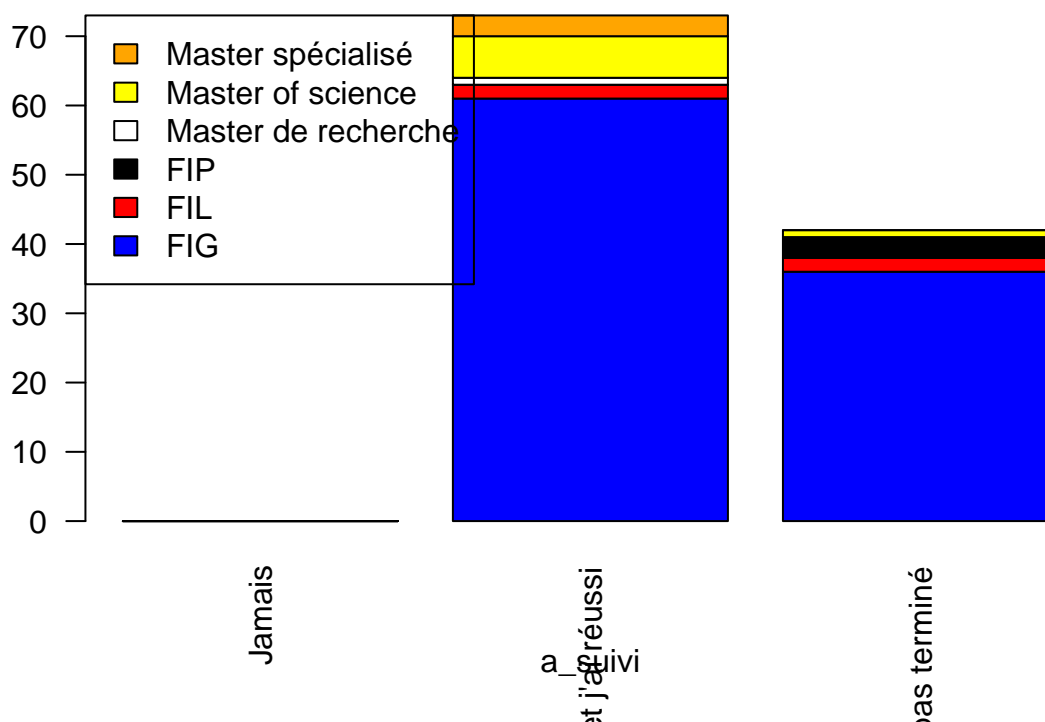
Distribution du temps alloué par cours vs motivation certificat



Les master recherche accordent plus du temps pour leurs moocs.

```
count1 <- table(data_suivi$formation, data_suivi$a_suivi)
barplot(count1, main="proportion terminé/non par formation", xlab="a_suivi", col=c("blue","red","black"))
```

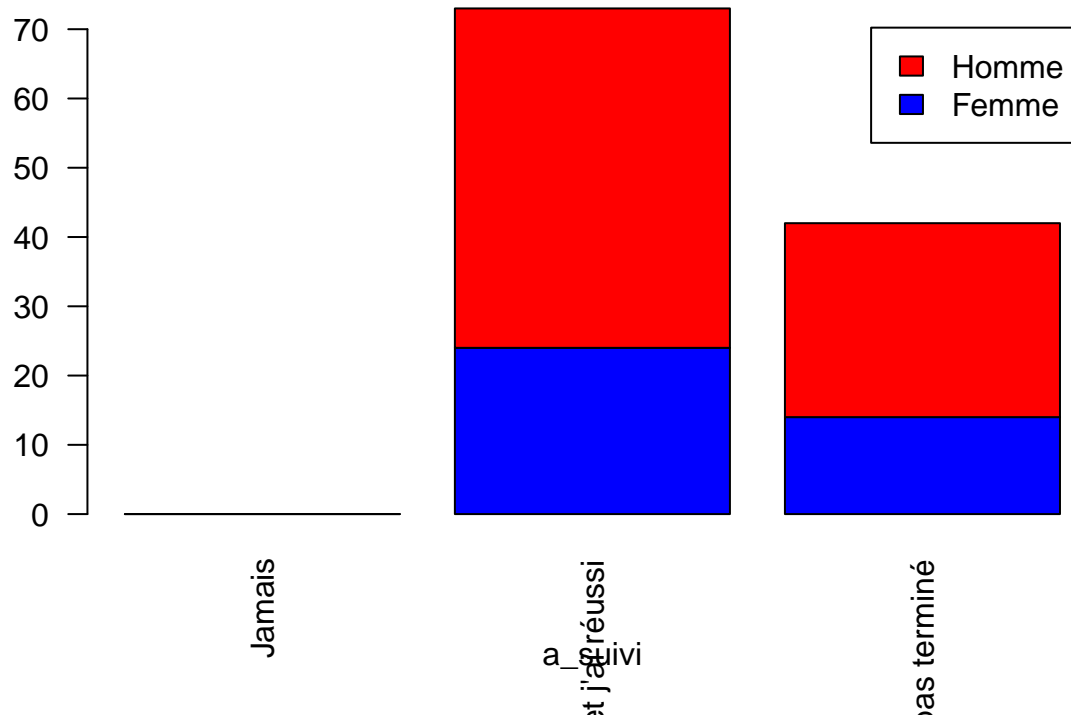
proportion terminé/non par formation



Seuls les FIG et FIL (nantes) qui ont n'ont pas fini leurs moocs.

```
count2 <- table(data_suivi$sexe, data_suivi$a_suivi)
barplot(count2, main="proportion terminé/non par sexe", xlab="a_suivi", col=c("blue","red"), legend = rownames(count2))
```

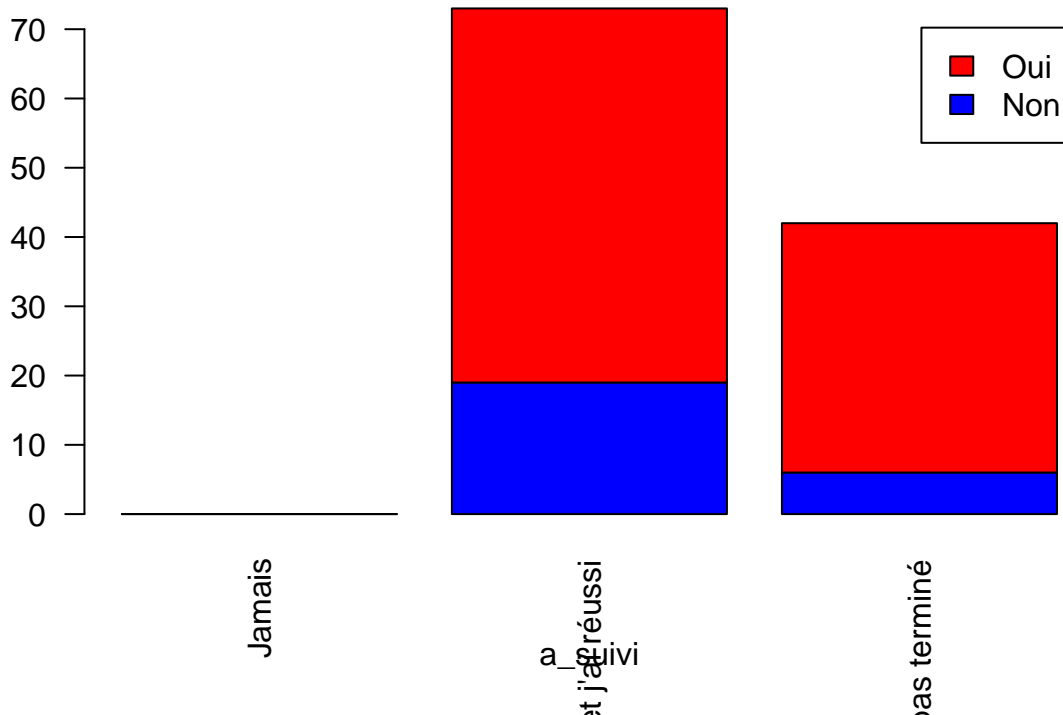
proportion terminé/non par sex



Les hommes sont les plus qui ne finissent pas leurs moocs.

```
count3 <- table(data_suivi$mooc_inclu_etude, data_suivi$a_suivi)
barplot(count3, main="proportion terminé/non par mooc_inclu_etude", xlab="a_suivi", col=c("blue","red"))
```


proportion terminé/non par mooc_inclu_etude



la majorité des personnes qui n'ont pas terminé ses moocs ils ont fait ses mooc pendant leur étude=> ce qui peut être naturelle car il y a une charge de travail pendant les études.

Test statistiques :

Hypothèse 1 : la terminaison des mooc ne dépend pas du genre :

Chi-square test basics :

Chi-square test examines whether rows and columns of a contingency table are statistically significantly associated.

Null hypothesis (H0): the row and the column variables of the contingency table are independent. Alternative hypothesis (H1): row and column variables are dependent

```
# contingency table
```

```
table(data_suivi$a_suivi,data_suivi$sexe)
```

```
##
##
##               Femme Homme
##  Jamais           0      0
##  Oui, je l'ai suivi et j'ai réussi      24      49
##  Oui, je l'ai suivi mais je ne l'ai pas terminé      14      28
```

```
chisq_sexe <- chisq.test(data_suivi$a_suivi,data_suivi$sexe)
```

```
chisq_sexe
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
```

```
##
## data: data_suivi$a_suivi and data_suivi$sexe
## X-squared = 0, df = 1, p-value = 1
```

le test donne une p-value de 1 ce qui permet avec un risque de 5% de garder l'hypothèse nulle et donc conclure que les deux variables catégorielles sont indépendentes.

Hypothèse 2 : la terminaison des mooc ne depend pas de la formation suivie :

```
# contingency table
```

```
table(data_suivi$a_suivi,data_suivi$formation)
```

```
##
##
##          FIG FIL FIP
##  Jamais          0  0  0
##  Oui, je l'ai suivi et j'ai réussi      61  2  0
##  Oui, je l'ai suivi mais je ne l'ai pas terminé 36  2  3
##
##          Master de recherche
##  Jamais                      0
##  Oui, je l'ai suivi et j'ai réussi      1
##  Oui, je l'ai suivi mais je ne l'ai pas terminé 0
##
##          Master of science
##  Jamais                      0
##  Oui, je l'ai suivi et j'ai réussi      6
##  Oui, je l'ai suivi mais je ne l'ai pas terminé 1
##
##          Master spécialisé
##  Jamais                      0
##  Oui, je l'ai suivi et j'ai réussi      3
##  Oui, je l'ai suivi mais je ne l'ai pas terminé 0
```

```
chisq_formation <- chisq.test(data_suivi$a_suivi,data_suivi$formation)
```

```
## Warning in chisq.test(data_suivi$a_suivi, data_suivi$formation): Chi-
## squared approximation may be incorrect
```

```
chisq_formation
```

```
##
## Pearson's Chi-squared test
##
## data: data_suivi$a_suivi and data_suivi$formation
## X-squared = 9.3367, df = 5, p-value = 0.09637
```

Avec un risque de 5% on garde H_0 , donc les deux variables sont indépendentes.

Il y a une forte association positive entre la colonne FIP et la rangée "Oui, je l'ai suivi..pas terminé"

Hypothèse 3 : la terminaison des mooc ne depend pas du background du répondant :

```
table(data_suivi$a_suivi,data_suivi$background)
```

```
##
##
##          Autre école DUT Prépa
##  Jamais          0  0  0
```

```
## Oui, je l'ai suivi et j'ai réussi 4 2 59
## Oui, je l'ai suivi mais je ne l'ai pas terminé 0 2 36
##
## Université
## Jamais 0
## Oui, je l'ai suivi et j'ai réussi 8
## Oui, je l'ai suivi mais je ne l'ai pas terminé 4
```

```
chisq_background <- chisq.test(data_suivi$a_suivi,data_suivi$background)
```

```
## Warning in chisq.test(data_suivi$a_suivi, data_suivi$background): Chi-
## squared approximation may be incorrect
```

```
chisq_background
```

```
##
## Pearson's Chi-squared test
##
## data: data_suivi$a_suivi and data_suivi$background
## X-squared = 2.7447, df = 3, p-value = 0.4327
```

Toujours avec un risque de 5% on accepte $H_0 \Rightarrow$ les colonnes et les lignes sont indépendents. ##### Hypothèse 4: la terminaison des mooc ne depend pas de la plateforme utilisée :

```
#table(data_suivi$a_suivi,data_suivi$plateformes_utilisees)
```

```
chisq_plateforme <- chisq.test(data_suivi$a_suivi,data_suivi$plateformes_utilisees)
```

```
## Warning in chisq.test(data_suivi$a_suivi, data_suivi
## $plateformes_utilisees): Chi-squared approximation may be incorrect
```

```
chisq_plateforme
```

```
##
## Pearson's Chi-squared test
##
## data: data_suivi$a_suivi and data_suivi$plateformes_utilisees
## X-squared = 28.772, df = 29, p-value = 0.477
```

même conclusion

Hypthèse 4: la terminaison des mooc ne depend pas de périodes de suivi des mooc :

```
table(data_suivi$a_suivi,data_suivi$vacances_etude)
```

```
##
## Les deux Les vacances
## Jamais 0 0
## Oui, je l'ai suivi et j'ai réussi 34 6
## Oui, je l'ai suivi mais je ne l'ai pas terminé 24 3
##
## Vos études
## Jamais 0
## Oui, je l'ai suivi et j'ai réussi 33
## Oui, je l'ai suivi mais je ne l'ai pas terminé 15
```

```
chisq_periode <- chisq.test(data_suivi$a_suivi,data_suivi$vacances_etude)
```

```
## Warning in chisq.test(data_suivi$a_suivi, data_suivi$vacances_etude): Chi-
```

```
## squared approximation may be incorrect
```

```
chisq_periode
```

```
##
```

```
## Pearson's Chi-squared test
```

```
##
```

```
## data: data_suivi$a_suivi and data_suivi$vacances_etude
```

```
## X-squared = 1.2052, df = 2, p-value = 0.5474
```

=====

On veut comparer le degré de satisfaction envers les moocs en fonction de la catégorie du répondeur à notre sondage : le sexe a-t-il une influence sur la satisfaction? Et qu'en est-il du niveau actuel du répondeur? De sa formation à l'école? Du son background?

Hypothèses 5 : Le degré de satisfaction des femmes et des hommes est le même :

```
satisf_homme_femme<-data_suivi[,c("satisfaction","sexe")]
```

```
str(satisf_homme_femme)
```

```
## 'data.frame': 115 obs. of 2 variables:
```

```
## $ satisfaction: num 7 10 9 6 1 7 3 6 7 6 ...
```

```
## $ sexe : Factor w/ 2 levels "Femme","Homme": 2 2 2 2 2 1 2 2 2 2 ...
```

```
library(plyr); library(dplyr)
```

```
## -----
```

```
## You have loaded plyr after dplyr - this is likely to cause problems.
```

```
## If you need functions from both plyr and dplyr, please load plyr first, then dplyr:
```

```
## library(plyr); library(dplyr)
```

```
## -----
```

```
##
```

```
## Attaching package: 'plyr'
```

```
## The following objects are masked from 'package:dplyr':
```

```
##
```

```
## arrange, count, desc, failwith, id, mutate, rename, summarise,
```

```
## summarise
```

```
res<-ddply(satisf_homme_femme,~sexe,summarise,mean=mean(satisfaction),sd=sd(satisfaction))
```

le resultat est le suivant :

sexe	mean	sd
Femme	6.578947	1.926088
Homme	6.610390	1.987830

Les variances ne sont pas égales, donc allons voir si les distributions sont normales ou non ? Pour les hommes :

```
homme_series<-filter(satisf_homme_femme,sexe=="Homme")
```

```
shapiro.test(homme_series$satisfaction)
```

```
##
```

```
## Shapiro-Wilk normality test
```

```
##
## data:  homme_series$satisfaction
## W = 0.93798, p-value = 0.0009714
```

La distribution de la satisfaction des hommes n'est pas normale. On procède de même pour les femmes :

```
femme_series<-filter(satisf_homme_femme,sexe=="Femme")
shapiro.test(femme_series$satisfaction)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  femme_series$satisfaction
## W = 0.92978, p-value = 0.01974
```

Même conclusion qu'avant. Donc on doit utiliser un test de Wilcoxon qui est non paramétriques:

```
wilcox.test(femme_series$satisfaction,homme_series$satisfaction)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  femme_series$satisfaction and homme_series$satisfaction
## W = 1449.5, p-value = 0.9374
## alternative hypothesis: true location shift is not equal to 0
```

Avec un risque de 5%, on accepte H0 et donc il n'y a pas de différence de satisfaction entre les hommes et les femmes.

Hypothèses 6: Le degré de satisfaction est le même pour tout les niveaux d'études :

```
satisf_niveau<-data_suivi[,c("satisfaction","niveau_actuel")]
str(satisf_niveau)
```

```
## 'data.frame':   115 obs. of  2 variables:
## $ satisfaction : num  7 10 9 6 1 7 3 6 7 6 ...
## $ niveau_actuel: Factor w/ 5 levels "BAC +2","BAC +3",...: 4 4 2 4 2 3 4 4 3 4 ...
res2<-ddply(satisf_niveau,~niveau_actuel,summarise,mean=mean(satisfaction),sd=sd(satisfaction))
```

Le resultat est le suivant :

niveau_actuel	mean	sd
BAC +2	6.000000	2.380476
BAC +3	5.450000	2.372540
BAC +4	7.027778	1.521017
BAC +5	6.866667	1.914063
BAC +6 ou plus	6.750000	1.892969

Les variances ne sont pas égales, il faut voir la distribution pour choisir le test convenable:

```
bac2<-filter(satisf_niveau,niveau_actuel=="BAC +2")
shapiro.test(bac2$satisfaction)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  bac2$satisfaction
## W = 0.96061, p-value = 0.8239
```

==> avec un risque de 5% la distribution est normale pour BAC +2

```
bac3<-filter(satisf_niveau,niveau_actuel=="BAC +3")
shapiro.test(bac3$satisfaction)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  bac3$satisfaction
## W = 0.93991, p-value = 0.2389

==> avec un risque de 5% Distribution est normale pour BAC+3

bac4<-filter(satisf_niveau,niveau_actuel=="BAC +4")
shapiro.test(bac4$satisfaction)

##
##  Shapiro-Wilk normality test
##
## data:  bac4$satisfaction
## W = 0.93636, p-value = 0.03917

==> avec un risque de 5% la distribution n'est pas normale pour BAC+4

bac5<-filter(satisf_niveau,niveau_actuel=="BAC +5")
shapiro.test(bac5$satisfaction)

##
##  Shapiro-Wilk normality test
##
## data:  bac5$satisfaction
## W = 0.94592, p-value = 0.03567

==> avec un risque de 5% non normale pour BAC+5

bac6<-filter(satisf_niveau,niveau_actuel=="BAC +6 ou plus")
shapiro.test(bac6$satisfaction)

##
##  Shapiro-Wilk normality test
##
## data:  bac6$satisfaction
## W = 0.79065, p-value = 0.08649

==> avec un risque de 5% ,normale pour BAC+6 Donc on va utiliser le test de wilcoxon : #####BAC+2
VS BAC+3 :

wilcox.test(bac2$satisfaction,bac3$satisfaction)

## Warning in wilcox.test.default(bac2$satisfaction, bac3$satisfaction):
## cannot compute exact p-value with ties

##
##  Wilcoxon rank sum test with continuity correction
##
## data:  bac2$satisfaction and bac3$satisfaction
## W = 76, p-value = 0.7589
## alternative hypothesis: true location shift is not equal to 0

avec un risque de 5%, on accepte H0==> les deux niveau ont le même degré de satisfaction. #####BAC+3
VS BAC+4 :

wilcox.test(bac3$satisfaction,bac4$satisfaction)

## Warning in wilcox.test.default(bac3$satisfaction, bac4$satisfaction):
## cannot compute exact p-value with ties
```

```
##
## Wilcoxon rank sum test with continuity correction
##
## data: bac3$satisfaction and bac4$satisfaction
## W = 222, p-value = 0.01681
## alternative hypothesis: true location shift is not equal to 0
```

Avec un risque de 5%, on rejete $H_0 \implies$ les deux niveau n'ont pas le même degré de satisfaction. Pour savoir qui est le plus satisfait, on fait le test suivant :

```
wilcox.test(bac3$satisfaction,bac4$satisfaction,alternative = "greater")
```

```
## Warning in wilcox.test.default(bac3$satisfaction, bac4$satisfaction,
## alternative = "greater"): cannot compute exact p-value with ties
```

```
##
## Wilcoxon rank sum test with continuity correction
##
## data: bac3$satisfaction and bac4$satisfaction
## W = 222, p-value = 0.992
## alternative hypothesis: true location shift is greater than 0
```

Avec un risque de 5%, on affirme que le degré de satisfaction des BAC+4 est plus grand à celui des BAC+3.
#####BAC+4 VS BAC+5 :

```
wilcox.test(bac4$satisfaction,bac5$satisfaction)
```

```
## Warning in wilcox.test.default(bac4$satisfaction, bac5$satisfaction):
## cannot compute exact p-value with ties
```

```
##
## Wilcoxon rank sum test with continuity correction
##
## data: bac4$satisfaction and bac5$satisfaction
## W = 848.5, p-value = 0.7129
## alternative hypothesis: true location shift is not equal to 0
```

avec un risque de 5%, on affirme que les deux niveau ont le même degré de satisfaction.

BAC+5 vS BAC+6 ou plus :

```
wilcox.test(bac5$satisfaction,bac6$satisfaction)
```

```
## Warning in wilcox.test.default(bac5$satisfaction, bac6$satisfaction):
## cannot compute exact p-value with ties
```

```
##
## Wilcoxon rank sum test with continuity correction
##
## data: bac5$satisfaction and bac6$satisfaction
## W = 88, p-value = 0.9555
## alternative hypothesis: true location shift is not equal to 0
```

avec un risque de 5%, on affirme que les deux niveau ont le même degré de satisfaction.

Hypothèses 6: Le degré de satisfaction ne dépend pas du background : ... pareil

ACF:en cours ...