# TP2

```
##############################################################################
## Statistiques avec R - Ensai - 1A
## Author : BdL
## Last update : 12/01/2017
##############################################################################


##############################################################################
## Préambule
##############################################################################


## Chargement des librairies
library(MASS)

## Configuration par défaut des graphiques
old.par<-par(no.readonly=TRUE)
```

## Exercice 2

```
##############################################################################
## Exercice 2
##############################################################################

## Question 1

# Chargement des données
scol<-read.csv("../data/stat_R_education.txt",sep="\t")
scol
```

```
##                        Education Age_Group Count
## 1  Did not complete high school     25-34  5416
## 2  Did not complete high school     35-44  5030
## 3  Did not complete high school     45-54  5777
## 4  Did not complete high school     55-64  7606
## 5  Did not complete high school       >64 13746
## 6         Completed high school     25-34 16431
## 7         Completed high school     35-44  1855
## 8         Completed high school     45-54  9435
## 9         Completed high school     55-64  8795
## 10        Completed high school       >64  7558
## 11            College,1-3 years     25-34  8555
## 12            College,1-3 years     35-44  5576
## 13            College,1-3 years     45-54  3124
## 14            College,1-3 years     55-64  2524
## 15            College,1-3 years       >64  2503
## 16      College,4 or more years     25-34  9771
## 17      College,4 or more years     35-44  7596
## 18      College,4 or more years     45-54  3904
## 19      College,4 or more years     55-64  3109
## 20      College,4 or more years       >64  2483
```

```
typeof(scol);class(scol)
```

```
## [1] "list"
## [1] "data.frame"
```

```r
# tableau de contingence
educ<-xtabs(Count~Education+Age_Group,data=scol)
educ
```

```
##                              Age_Group
## Education                     25-34 35-44 45-54 55-64   >64
##   College,1-3 years            8555  5576  3124  2524  2503
##   College,4 or more years      9771  7596  3904  3109  2483
##   Completed high school       16431  1855  9435  8795  7558
##   Did not complete high school  5416  5030  5777  7606 13746
```

```r
rm(educ)

# Question 2
# changer le nom
names(scol)[match("Age_Group",names(scol))]<-"Age"

# education est bien un facteur
is.factor(scol$Age)
```

```
## [1] TRUE
```

```r
scol$Age<-factor(scol$Age,levels=c("25-34","35-44","45-54","55-64",">64"),ordered=TRUE)

# regarder dans quel sens sont rangés les niveaux
levels(scol$Education)
```

```
## [1] "College,1-3 years"            "College,4 or more years"
## [3] "Completed high school"        "Did not complete high school"
```

```r
# affectation des facteurs
levels(scol$Education)<-c("1-3","4+","0","-1")
scol
```

```
##    Education   Age Count
## 1         -1 25-34  5416
## 2         -1 35-44  5030
## 3         -1 45-54  5777
## 4         -1 55-64  7606
## 5         -1   >64 13746
## 6          0 25-34 16431
## 7          0 35-44  1855
## 8          0 45-54  9435
## 9          0 55-64  8795
## 10         0   >64  7558
## 11       1-3 25-34  8555
## 12       1-3 35-44  5576
## 13       1-3 45-54  3124
## 14       1-3 55-64  2524
## 15       1-3   >64  2503
## 16        4+ 25-34  9771
## 17        4+ 35-44  7596
## 18        4+ 45-54  3904
## 19        4+ 55-64  3109
## 20        4+   >64  2483
```

```r
# ordonancement des niveaux de facteurs
scol$Education<-factor(scol$Education,levels=c("-1","0","1-3","4+"),ordered=TRUE)
scol$Education
```

```
##  [1] -1  -1  -1  -1  -1  0   0   0   0   0   1-3 1-3 1-3 1-3 1-3 4+  4+
## [18] 4+  4+  4+
## Levels: -1 < 0 < 1-3 < 4+
```

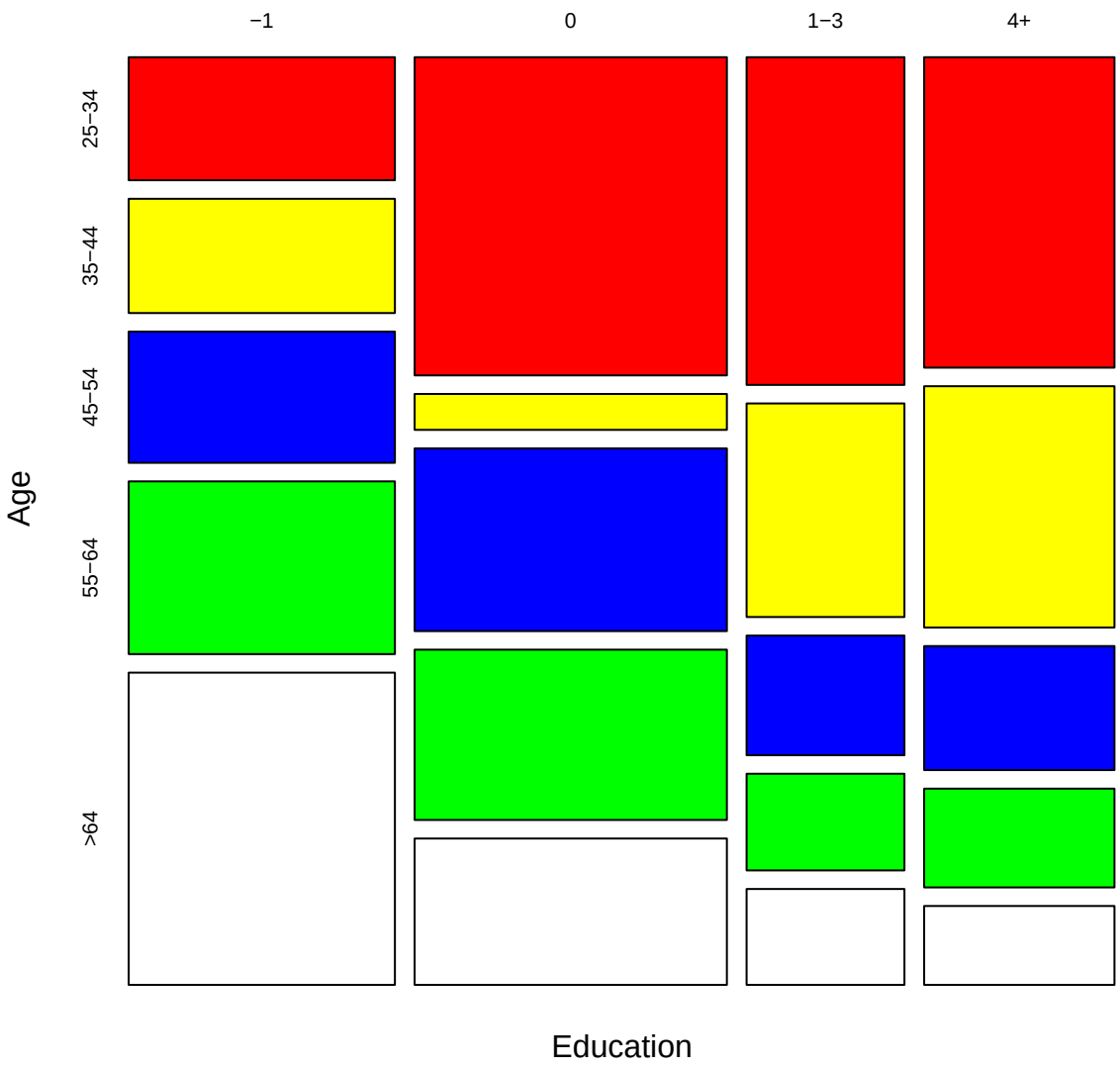
```r
# tableau de contingence
cont<-xtabs(Count ~ Education + Age, data=scol)
cont
```

```
##           Age
## Education 25-34 35-44 45-54 55-64   >64
##       -1   5416  5030  5777  7606 13746
##       0   16431  1855  9435  8795  7558
##       1-3  8555  5576  3124  2524  2503
##       4+   9771  7596  3904  3109  2483
```

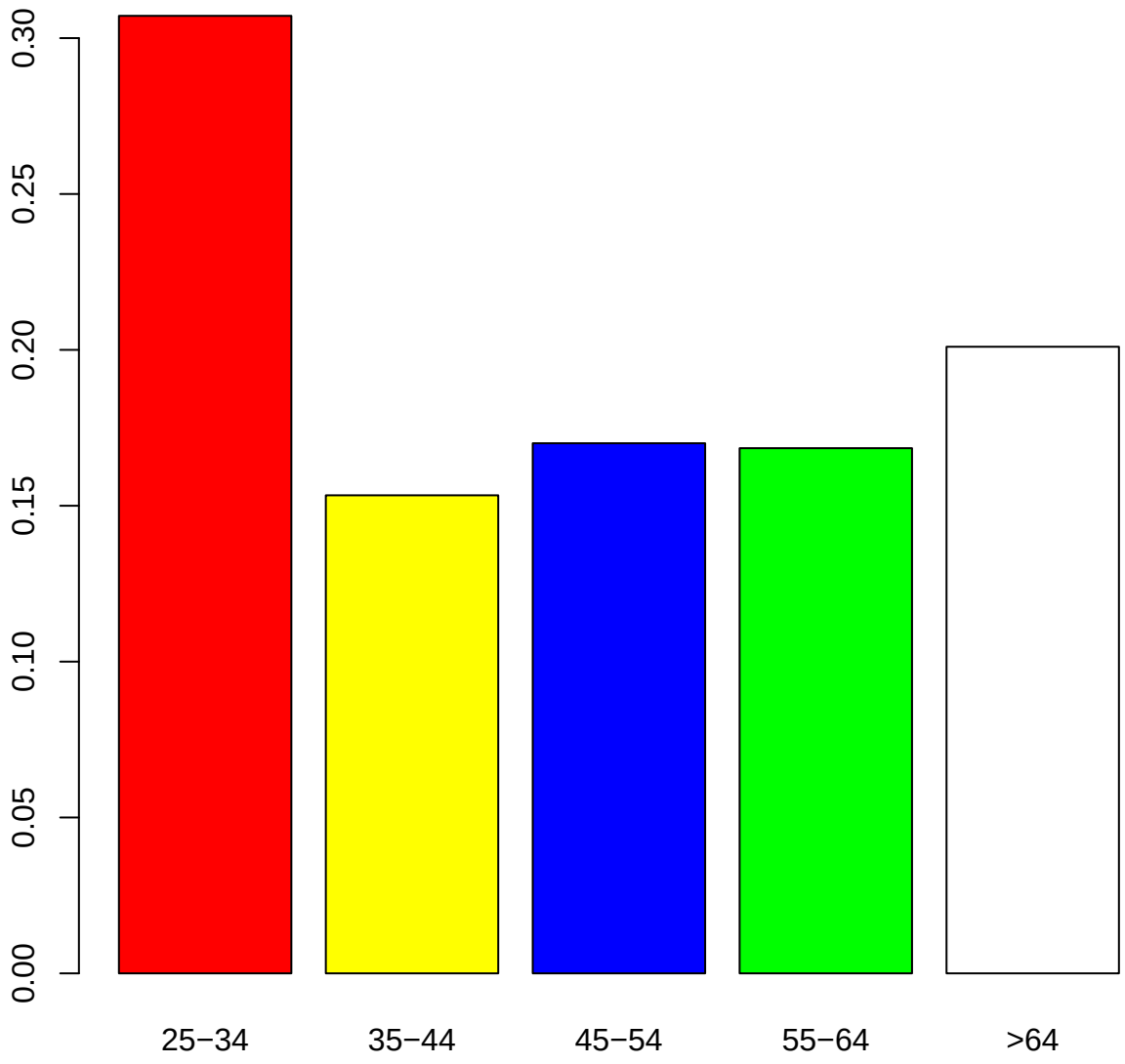
```r
str(cont)
```

```
##  xtabs [1:4, 1:5] 5416 16431 8555 9771 5030 ...
##  - attr(*, "dimnames")=List of 2
##   ..$ Education: chr [1:4] "-1" "0" "1-3" "4+"
##   ..$ Age      : chr [1:5] "25-34" "35-44" "45-54" "55-64" ...
##  - attr(*, "class")= chr [1:2] "xtabs" "table"
##  - attr(*, "call")= language xtabs(formula = Count ~ Education + Age, data = scol)
```

```r
plot(cont,col=c("red","yellow","blue","green","white"))
```

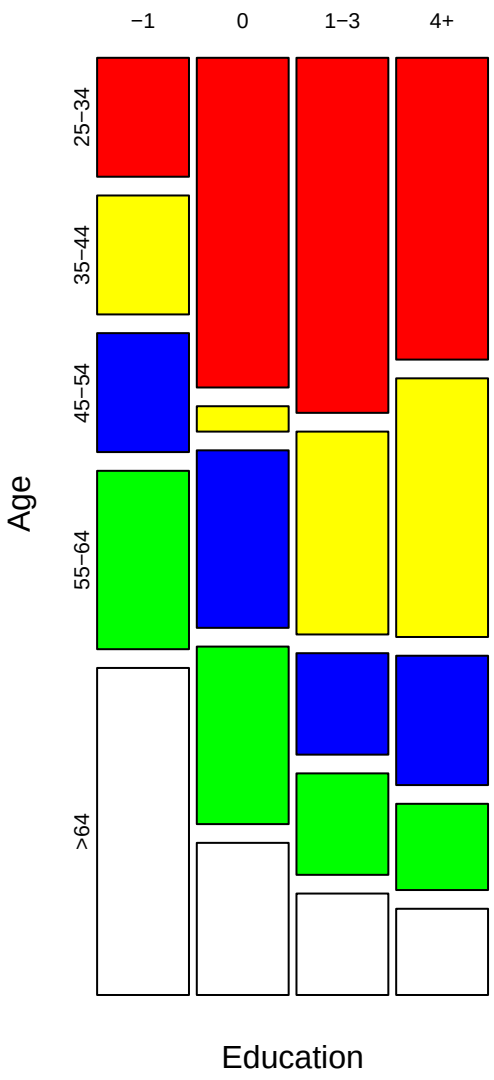


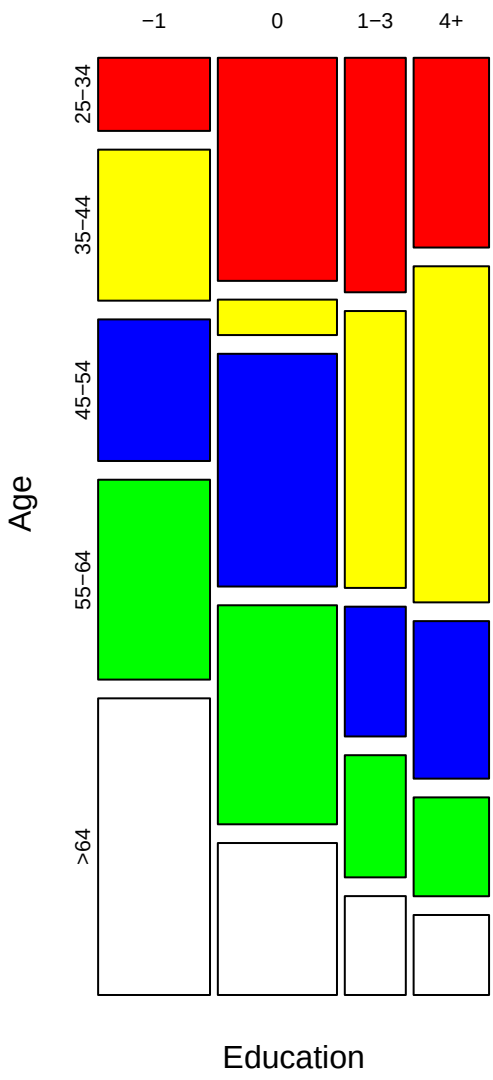

```
## Question 3 :
## margin.table
mar_age<-margin.table(cont,2)

## manuellement
mmar_age<-apply(cont,2,sum)

## barplot
barplot(mar_age/sum(mar_age),col=c("red","yellow","blue","green","white"))
```




```
## Question 4
total<-summary(cont)$n.cases
total

## [1] 130794

freq_cont<-round(cont/total,digits=2)
freq_cont

##           Age
## Education 25-34 35-44 45-54 55-64  >64
##       -1   0.04  0.04  0.04  0.06 0.11
##       0    0.13  0.01  0.07  0.07 0.06
##       1-3  0.07  0.04  0.02  0.02 0.02
##       4+   0.07  0.06  0.03  0.02 0.02

## Question 5
freq_marg_age<-margin.table(freq_cont,2)
freq_marg_age
```

```
## Age
## 25-34 35-44 45-54 55-64   >64
##  0.31  0.15  0.16  0.17  0.21

freq_marg_edu<-margin.table(freq_cont,1)
freq_marg_edu

## Education
##   -1    0  1-3   4+
## 0.29 0.34 0.17 0.20

# sous l'hypothese d'indépendance des variables le tableau celui des produits
# des frequences marginales
ind_cont<-as.table(freq_marg_edu%*%t(freq_marg_age))

summary(freq_cont)

## Call: xtabs(formula = Count ~ Education + Age, data = scol)
## Number of cases in table: 1
## Number of factors: 2
## Test for independence of all factors:
##  Chisq = 0.19226, df = 12, p-value = 1
##  Chi-squared approximation may be incorrect

summary(ind_cont)

## Number of cases in table: 1
## Number of factors: 2
## Test for independence of all factors:
##  Chisq = 1.937e-33, df = 12, p-value = 1
##  Chi-squared approximation may be incorrect

## Question 6
## loi conditionnelle de age sachant facteur Education
mult<-matrix(rep(freq_marg_edu,length(freq_marg_age)),nrow=length(freq_marg_edu))
age_wrt_Education<-(mult^(-1))*freq_cont
age_wrt_Education

##           Age
## Education      25-34      35-44      45-54      55-64        >64
##       -1  0.13793103 0.13793103 0.13793103 0.20689655 0.37931034
##       0   0.38235294 0.02941176 0.20588235 0.20588235 0.17647059
##       1-3 0.41176471 0.23529412 0.11764706 0.11764706 0.11764706
##       4+  0.35000000 0.30000000 0.15000000 0.10000000 0.10000000

## loi conditionnelle de Education sachant age
mult<-matrix(rep(freq_marg_age,length(freq_marg_edu)),byrow=TRUE,ncol=length(freq_marg_age))
Education_wrt_age<-(mult^(-1))*freq_cont
Education_wrt_age

##           Age
## Education      25-34      35-44      45-54      55-64        >64
##       -1  0.12903226 0.26666667 0.25000000 0.35294118 0.52380952
##       0   0.41935484 0.06666667 0.43750000 0.41176471 0.28571429
##       1-3 0.22580645 0.26666667 0.12500000 0.11764706 0.09523810
##       4+  0.22580645 0.40000000 0.18750000 0.11764706 0.09523810

## diagrammes en mosaique
par(mfrow=c(1,2))
plot(age_wrt_Education,col=c("red","yellow","blue","green","white"))
plot(Education_wrt_age,col=c("red","yellow","blue","green","white"))
```


age_wrt_Education
−1
0
1−3
4+
25−34
35−44
45−54
55−64
>64
Age
Education


Education_wrt_age
−1
0
1−3
4+
25−34
35−44
45−54
55−64
>64
Age
Education