TP4: Analyse Factorielle des Correspondances simple - Correction

```
rm(list=ls())
library("tidyverse")
```

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

## v ggplot2 3.3.3     v purrr   0.3.4
## v tibble  3.0.6     v dplyr   1.0.4
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library("ggplot2") #pour avoir de 'beaux' graphiques
library("FactoMineR") #pour effectuer l'ACP
library("factoextra") #pour extraire et visualiser les résultats issus de FactoMineR
```

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

```
library("questionr")
library("scales")
```

```
##
## Attaching package: 'scales'

## The following object is masked from 'package:purrr':
##
##     discard

## The following object is masked from 'package:readr':
##
##     col_factor
```

```
library("ca") # pour avoir le jeu de données smoke
```

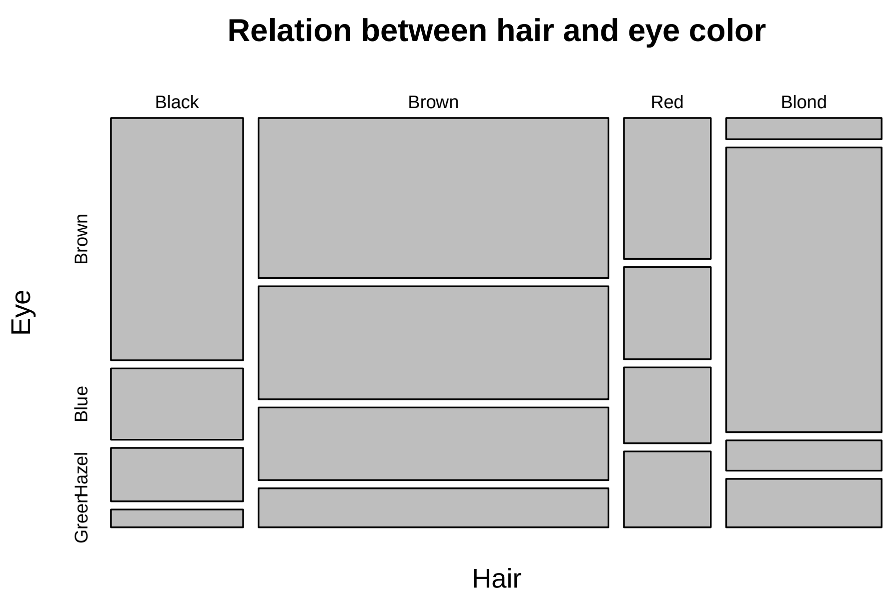
1 Couleur des yeux / Couleur des cheveux

1.1 Importation du jeu de données

```r
data(HairEyeColor)
HEC = HairEyeColor[,,1] +HairEyeColor[,,2]
knitr::kable(HEC)
```

	Brown	Blue	Hazel	Green
Black	68	20	15	5
Brown	119	84	54	29
Red	26	17	14	14
Blond	7	94	10	16

```r
mosaicplot(HEC, main = "Relation between hair and eye color")
```




```r
n = sum(HEC)
```

1.2 Fréquence marginale

```r
P=HEC/n
#lignes
rowSums(P)
```

```
##     Black     Brown       Red     Blond
## 0.1824324 0.4831081 0.1199324 0.2145270
```

```r
#colonne
colSums(P)
```

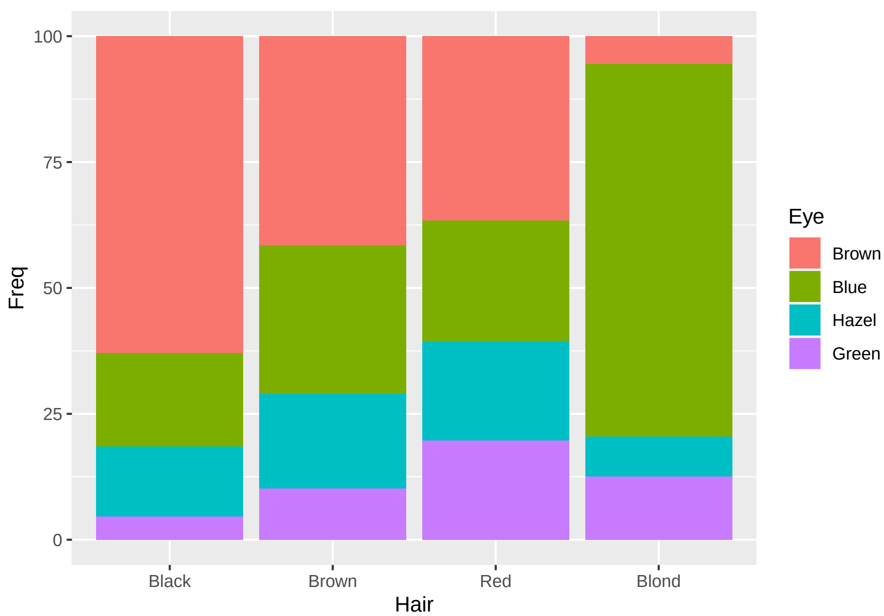
```
##     Brown      Blue     Hazel     Green
## 0.3716216 0.3631757 0.1570946 0.1081081
```

1.3 Profils lignes et colonnes

```
# Profil ligne
Profil.Ligne <- rprop(HEC,digits = 2)
knitr::kable(Profil.Ligne)
```

	Brown	Blue	Hazel	Green	Total
Black	62.962963	18.51852	13.888889	4.62963	100
Brown	41.608392	29.37063	18.881119	10.13986	100
Red	36.619718	23.94366	19.718310	19.71831	100
Blond	5.511811	74.01575	7.874016	12.59843	100
Ensemble	37.162162	36.31757	15.709459	10.81081	100

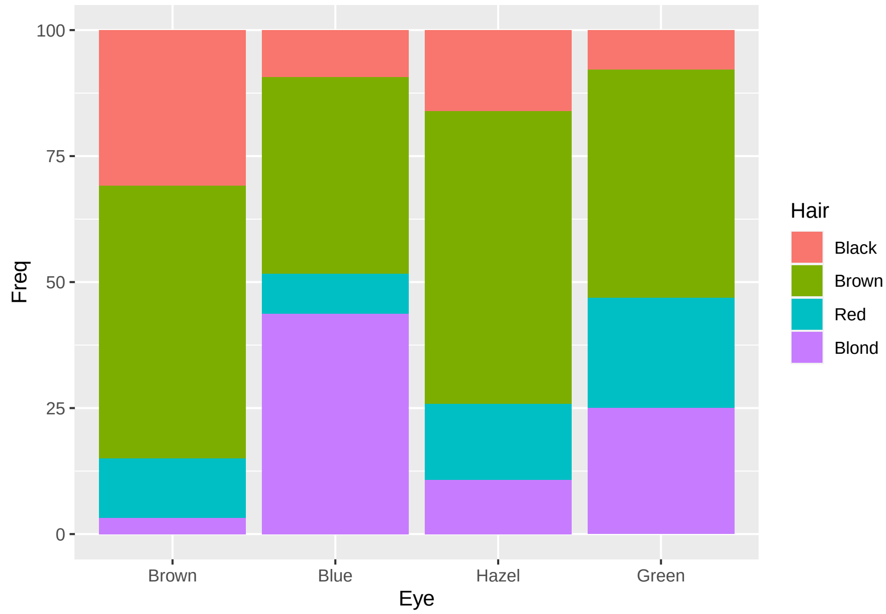
```
row<-rprop(HEC,total=FALSE) %>% as.data.frame()
row %>% ggplot(aes(x=Hair,fill=Eye))+geom_bar(aes(y=Freq), stat="identity")
```




```
# Profil colonne
Profil.Colonne <- cprop(HEC,digits = 2)
knitr::kable(Profil.Colonne)
```

	Brown	Blue	Hazel	Green	Ensemble
Black	30.909091	9.302326	16.12903	7.8125	18.24324
Brown	54.090909	39.069767	58.06452	45.3125	48.31081
Red	11.818182	7.906977	15.05376	21.8750	11.99324
Blond	3.181818	43.720930	10.75269	25.0000	21.45270
Total	100.000000	100.000000	100.00000	100.0000	100.00000

```r
col <- cprop(HEC,total=FALSE) %>% as.data.frame()
col %>% ggplot(aes(x=Eye,fill=Hair))+geom_bar(aes(y=Freq), stat="identity")
```



1.4 AFC

```r
res.ca <- CA(HEC,graph = FALSE)
```

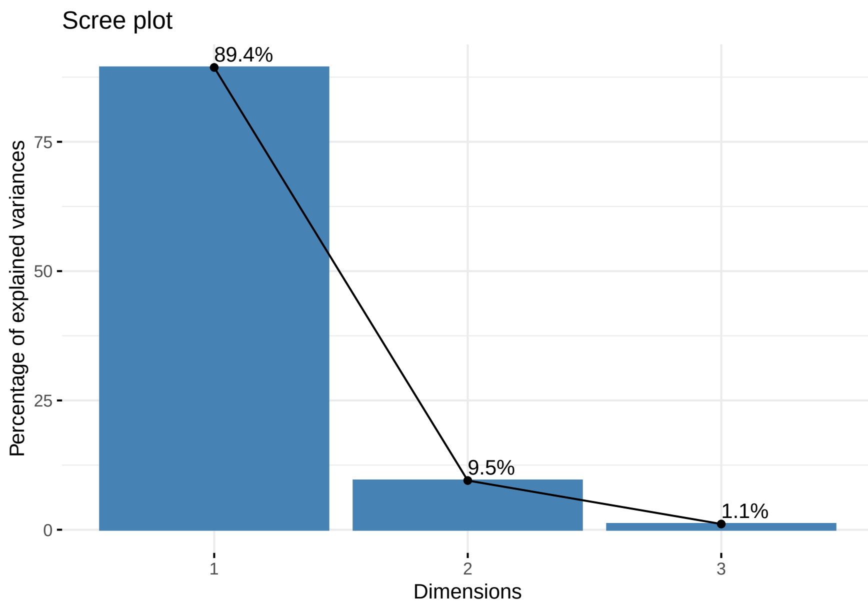
1.4.1 Valeurs propres et choix du nombre d'axes

Valeurs propres

```r
knitr::kable(res.ca$eig)
```

	eigenvalue	percentage of variance	cumulative percentage of variance
dim 1	0.2087727	89.372732	89.37273
dim 2	0.0222266	9.514911	98.88764
dim 3	0.0025984	1.112356	100.00000

```r
fviz_eig(res.ca, addlabels = TRUE)
```



Statistique de test du χ^2 ou Φ

```
h.chi <-  chisq.test(HEC)
phi <- h.chi$statistic/n
sum(res.ca$eig[,1])
```

```
## [1] 0.2335977
```

1.4.2 Qualité de représentation

```
rowSums(res.ca$row$cos2[,1:2]) #ligne
```

```
##     Black     Brown       Red     Blond
## 0.9898580 0.9063291 0.9450657 0.9996260
```

```
rowSums(res.ca$col$cos2[,1:2]) #colonne
```

```
##     Brown      Blue     Hazel     Green
## 0.9981354 0.9999277 0.8787352 0.9484270
```

1.4.3 Contributions

```
res.ca$row$contrib[,1:2] #ligne
```

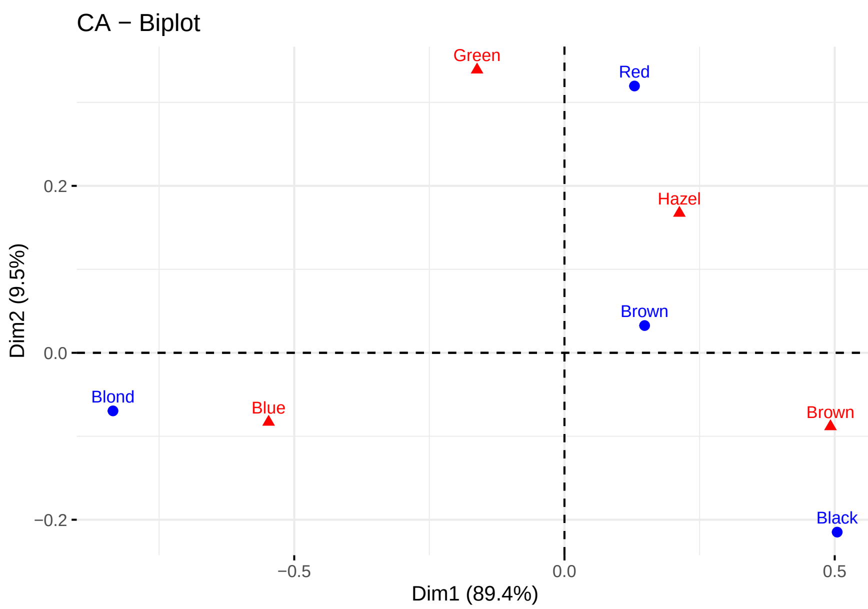
```
##            Dim 1     Dim 2
## Black 22.2463241 37.877386
## Brown  5.0859953  2.319381
## Red    0.9637371 55.130519
## Blond 71.7039435  4.672715
```

```
res.ca$col$contrib[,1:2] #colonne
```

```
##           Dim 1    Dim 2
## Brown 43.115744 13.04249
## Blue  52.128445 11.24401
## Hazel  3.400961 19.80398
## Green  1.354851 55.90952
```

1.4.4 Graphe de l'AFC

```
fviz_ca_biplot(res.ca, pointsize = 2, labelsize = 3)
```




```
#plot(res.ca, autoLab = "yes")
# bleu: profils-lignes
# rouge: profils-colonnes
```

1.4.5 Taux de liaison

```
P = HEC/sum(HEC)
DI <- apply(P, 1, sum)
DJ <- apply(P, 2, sum)
T = (P-t(t(DI))%*%DJ)/(t(t(DI))%*%DJ)
knitr::kable(T,digits=2)
```

	Brown	Blue	Hazel	Green
Black	0.69	-0.49	-0.12	-0.57
Brown	0.12	-0.19	0.20	-0.06
Red	-0.01	-0.34	0.26	0.82
Blond	-0.85	1.04	-0.50	0.17

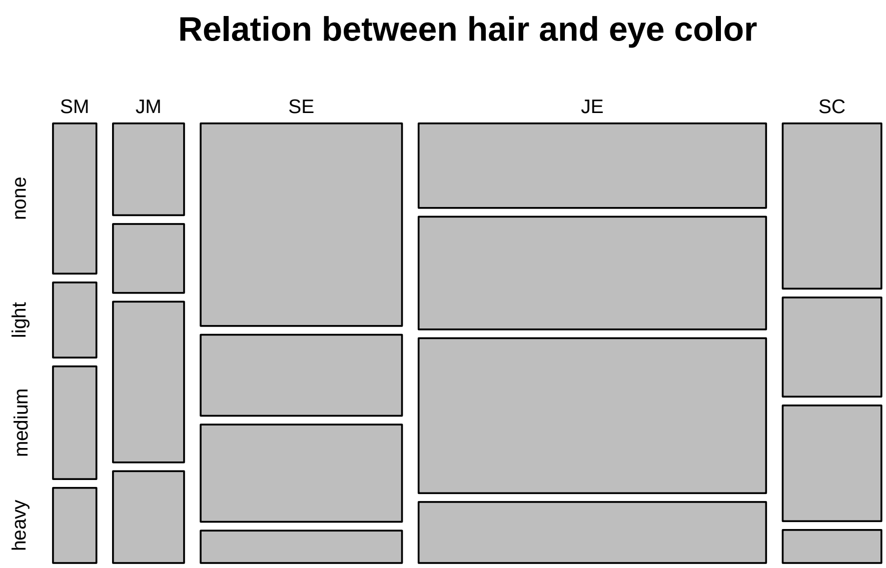
2 Fumeur

2.1 Importation du jeu de données

```r
data(smoke)
knitr::kable(smoke)
```

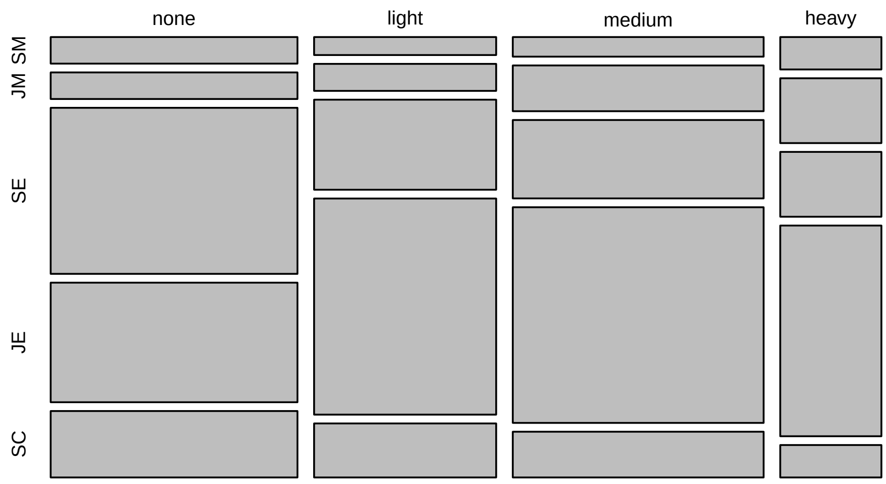
	none	light	medium	heavy
SM	4	2	3	2
JM	4	3	7	4
SE	25	10	12	4
JE	18	24	33	13
SC	10	6	7	2

```r
mosaicplot(smoke, main = "Relation between hair and eye color")
```




```r
mosaicplot(t(smoke), main = "Relation between hair and eye color")
```

Relation between hair and eye color




```
n = sum(smoke)
```

2.2 Test du χ^2

```
h.chi <-  chisq.test(smoke)
```

```
## Warning in chisq.test(smoke): Chi-squared approximation may be incorrect
```

```
h.chi
```

```
##
##  Pearson's Chi-squared test
##
## data:  smoke
## X-squared = 16.442, df = 12, p-value = 0.1718
```

2.3 AFC

```
res.ca <- CA(smoke,graph = FALSE)
```

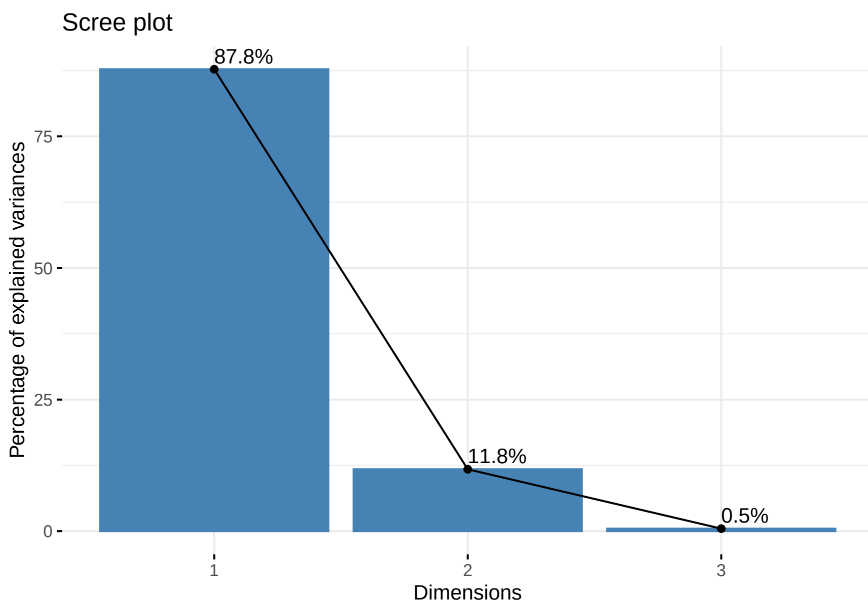
2.3.1 Valeurs propres et choix du nombre d'axes

Valeurs propres

```
knitr::kable(res.ca$eig)
```

	eigenvalue	percentage of variance	cumulative percentage of variance
dim 1	0.0747591	87.7558731	87.75587
dim 2	0.0100172	11.7586535	99.51453
dim 3	0.0004136	0.4854734	100.00000

```
fviz_eig(res.ca, addlabels = TRUE)
```



2.3.2 Coordonnées

```
res.ca$row$coord[,1:2] #ligne
```

```
##          Dim 1       Dim 2
## SM -0.06576838  0.19373700
## JM  0.25895842  0.24330457
## SE -0.38059489  0.01065991
## JE  0.23295191 -0.05774391
## SC -0.20108912 -0.07891123
```

```
res.ca$col$coord[,1:2] #colonne
```

```
##              Dim 1        Dim 2
## none   -0.39330845  0.030492071
## light   0.09945592 -0.141064289
## medium  0.19632096 -0.007359109
## heavy   0.29377599  0.197765656
```

2.3.3 Qualité de représentation

```r
res.ca$row$cos2[,1:2] #ligne
```

```
##         Dim 1        Dim 2
## SM 0.09223203 0.8003363898
## JM 0.52639991 0.4646824609
## SE 0.99903295 0.0007837197
## JE 0.94193412 0.0578762422
## SC 0.86534551 0.1332569974
```

```r
res.ca$col$cos2[,1:2] #colonne
```

```
##            Dim 1       Dim 2
## none   0.9940204 0.005974514
## light  0.3267262 0.657289655
## medium 0.9818480 0.001379626
## heavy  0.6843977 0.310154247
```

2.3.4 Contributions

```r
res.ca$row$contrib[,1:2] #ligne
```

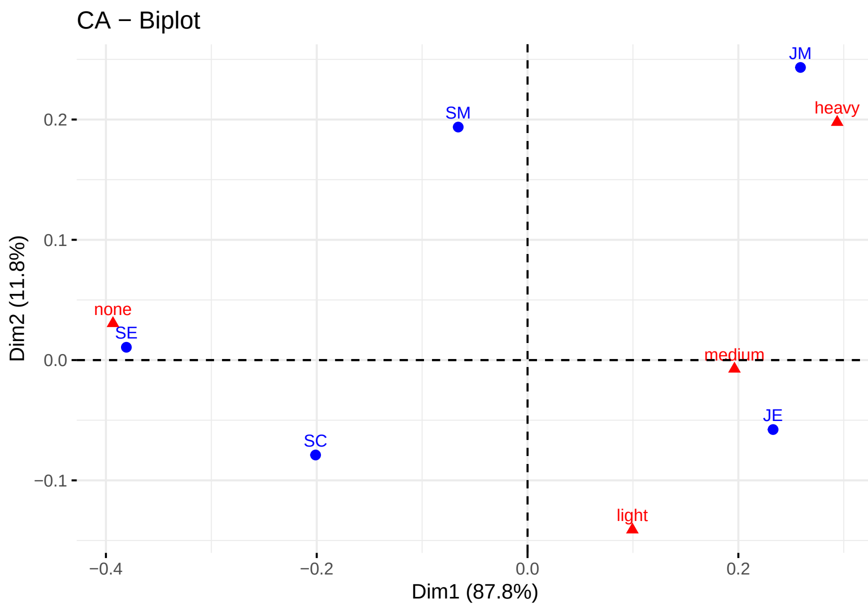
```
##         Dim 1      Dim 2
## SM  0.3297658 21.3557601
## JM  8.3658712 55.1150552
## SE 51.2005549  0.2997604
## JE 33.0973947 15.1772191
## SC  7.0064134  8.0522052
```

```r
res.ca$col$contrib[,1:2] #colonne
```

```
##           Dim 1      Dim 2
## none   65.39958  2.9335998
## light   3.08498 46.3173682
## medium 16.56165  0.1736758
## heavy  14.95379 50.5753561
```

2.3.5 Graphe de l’AFC

```r
fviz_ca_biplot(res.ca, pointsize = 2, labelsize = 3)
```




```r
#plot(res.ca, autoLab = "yes")
# bleu: profils-lignes
# rouge: profils-colonnes
```

2.4 Taux de liaison

```r
P = smoke/sum(smoke)
DI <- apply(P, 1, sum)
DJ <- apply(P, 2, sum)
T = (P-t(t(DI))%*%DJ)/(t(t(DI))%*%DJ)
knitr::kable(T,digits=2)
```

	none	light	medium	heavy
SM	0.15	-0.22	-0.15	0.40
JM	-0.30	-0.29	0.21	0.72
SE	0.55	-0.16	-0.27	-0.39
JE	-0.35	0.17	0.17	0.14
SC	0.27	0.03	-0.13	-0.38

3 Age & Loisirs

3.1 Importation du jeu de données

```r
LoisirsAge <- read.table("TV.txt",header=TRUE,sep="",row.names =1)
knitr::kable(LoisirsAge)
```

	Moins15	X15.24	X25.39	X40.60	Plus60
TV	322	114	72	135	130
Theatre	1	17	85	92	14
Cinema	90	220	192	87	7
Lecture	23	38	57	73	80
Restaurant	7	53	158	49	13
Night-Club	0	87	109	21	0
Concert	27	153	130	47	1

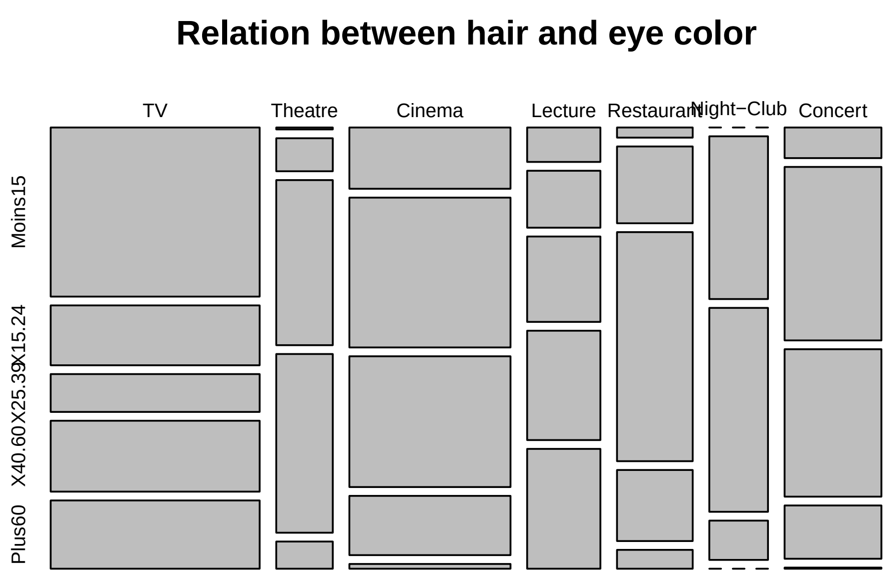
```
n = sum(LoisirsAge)
```

3.2 Fréquence marginale

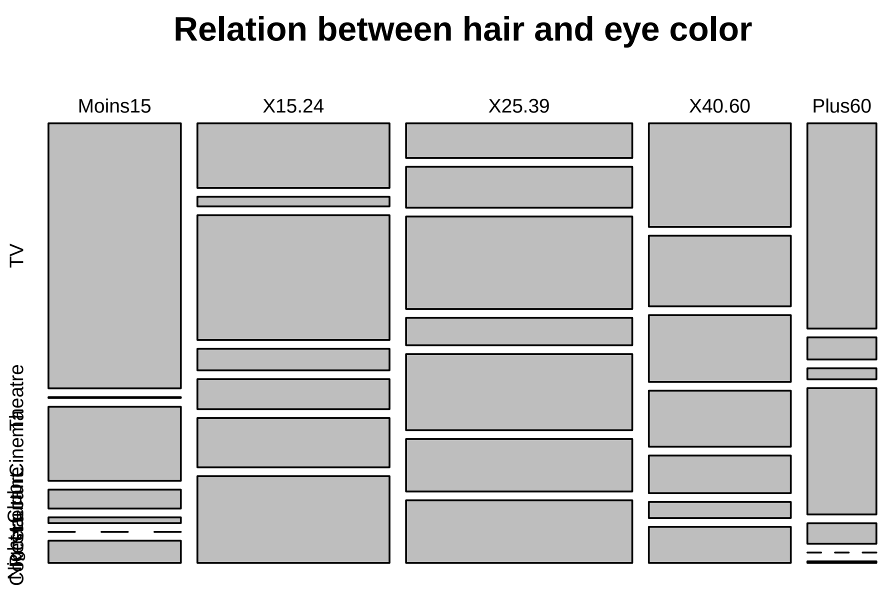
```
P=LoisirsAge/n
#lignes
Ml <- rowSums(P)
#colonne
Mc <- colSums(P)
```

3.3 Graphes mosaic

```
mosaicplot(LoisirsAge, main = "Relation between hair and eye color")
```




```
mosaicplot(t(LoisirsAge), main = "Relation between hair and eye color")
```



3.4 Test du χ^2

```
h.chi <-  chisq.test(LoisirsAge)
h.chi
```

```
##
##  Pearson's Chi-squared test
##
## data:  LoisirsAge
## X-squared = 1180.9, df = 24, p-value < 2.2e-16
```

3.5 AFC

```
res.ca <- CA(LoisirsAge,graph = FALSE)
```

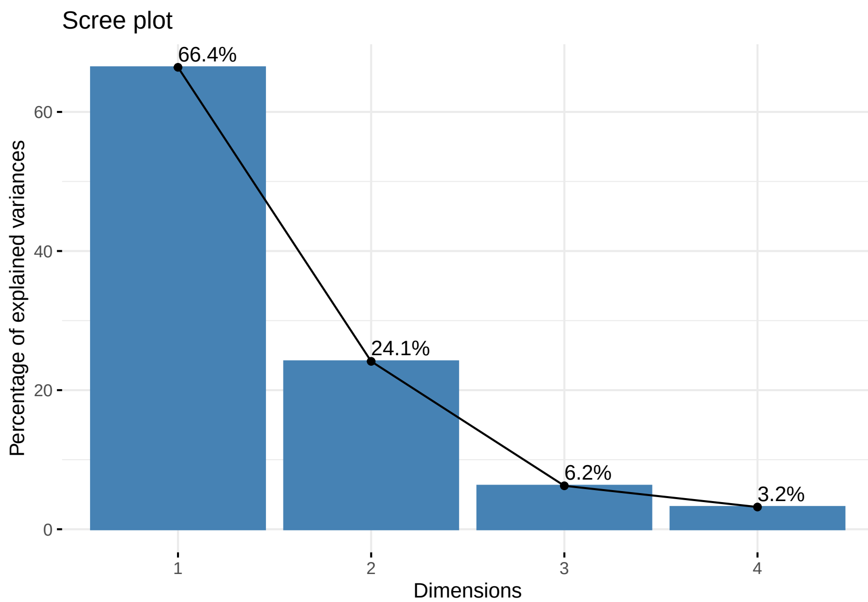
3.5.1 Valeurs propres et choix du nombre d'axes

Valeurs propres

```
knitr::kable(res.ca$eig)
```

	eigenvalue	percentage of variance	cumulative percentage of variance
dim 1	0.2900476	66.413392	66.41339
dim 2	0.1054239	24.139338	90.55273
dim 3	0.0272909	6.248906	96.80164
dim 4	0.0139682	3.198364	100.00000

```r
fviz_eig(res.ca, addlabels = TRUE)
```



3.5.2 Coordonnées

```r
res.ca$row$coord[,1:2] #ligne
```

```
##                 Dim 1      Dim 2
## TV          0.7508626 -0.1436471
## Theatre    -0.2618918  0.6669141
## Cinema     -0.2619314 -0.2505704
## Lecture     0.3493731  0.6047043
## Restaurant -0.5106749  0.2426317
## Night-Club -0.7069354 -0.1208485
## Concert    -0.4688726 -0.2362948
```

```r
res.ca$col$coord[,1:2] #colonne
```

```
##               Dim 1      Dim 2
## Moins15  0.82862049 -0.3860445
## X15.24  -0.33626403 -0.3205959
## X25.39  -0.50240298  0.1041274
## X40.60   0.06661964  0.3822494
## Plus60   0.85605603  0.5053850
```

3.5.3 Qualité de représentation

```r
res.ca$row$cos2[,1:2] #ligne
```

```
##                 Dim 1      Dim 2
## TV         0.9563292 0.03500098
## Theatre    0.1019969 0.66142848
## Cinema     0.5018104 0.45922330
## Lecture    0.1970449 0.59029930
## Restaurant 0.6423672 0.14500722
## Night-Club 0.9303677 0.02718803
## Concert    0.7490889 0.19025311
```

```r
res.ca$col$cos2[,1:2] #colonne
```

```
##              Dim 1      Dim 2
## Moins15 0.7990656 0.17343863
## X15.24  0.4521298 0.41097764
## X25.39  0.8881979 0.03815355
## X40.60  0.0216301 0.71211080
## Plus60  0.6531461 0.22764106
```

3.5.4 Contributions

```r
res.ca$row$contrib[,1:2] #ligne
```

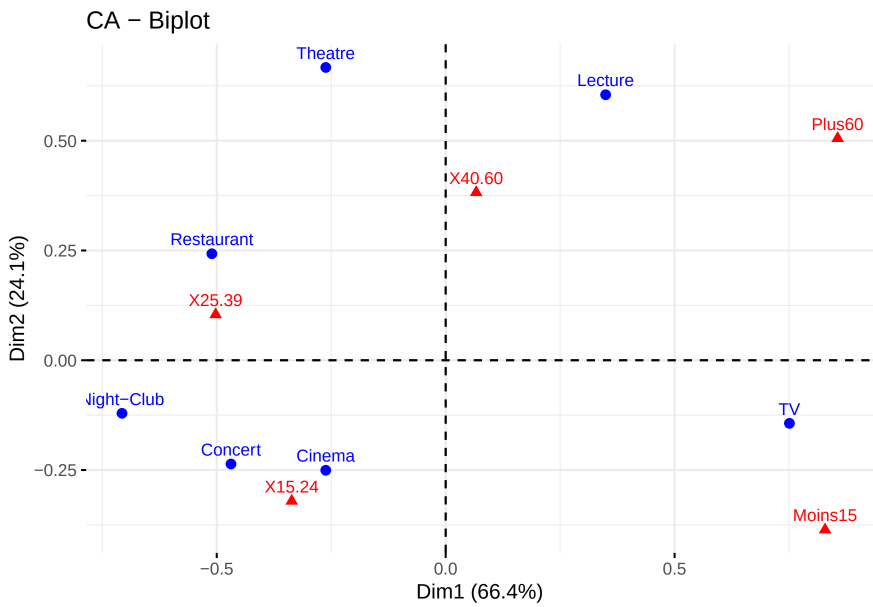
```
##                 Dim 1     Dim 2
## TV         55.567963  5.595355
## Theatre     1.827740 32.609225
## Cinema      5.213695 13.126838
## Lecture     4.217668 34.762406
## Restaurant  9.310459  5.782394
## Night-Club 13.827487  1.111723
## Concert    10.034988  7.012060
```

```r
res.ca$col$contrib[,1:2] #colonne
```

```
##               Dim 1     Dim 2
## Moins15 41.1465363 24.571234
## X15.24   9.8326207 24.589779
## X25.39  25.8430663  3.054215
## X40.60   0.2852064 25.833225
## Plus60  22.8925703 21.951547
```

3.5.5 Graphe de l'AFC

```r
fviz_ca_biplot(res.ca, pointsize = 2, labelsize = 3)
```




```
#plot(res.ca, autoLab = "yes")
# bleu: profils-lignes
# rouge: profils-colonnes
```

3.6 Taux de liaison

```
T = (P-t(t(Ml))%*%Mc)/(t(t(Ml))%*%Mc)
knitr::kable(T,digits=2)
```

	Moins15	X15.24	X25.39	X40.60	Plus60
TV	1.40	-0.42	-0.69	-0.06	0.86
Theatre	-0.97	-0.68	0.37	1.36	-0.26
Cinema	-0.13	0.46	0.08	-0.22	-0.87
Lecture	-0.51	-0.44	-0.29	0.45	2.26
Restaurant	-0.86	-0.25	0.90	-0.06	-0.49
Night-Club	-1.00	0.59	0.69	-0.48	-1.00
Concert	-0.57	0.69	0.22	-0.30	-0.97