

# Partie\_2\_Analyse

KABASSINA Gnimdou Ange & Ibrahima GNING

2022-06-27

## Contents

<b>Introduction</b>	<b>2</b>
<b>I. Caractéristiques sociodémographiques</b>	<b>3</b>
I.1 Sexe . . . . .	3
I.2 Situation matrimoniale . . . . .	4
I.3 Age . . . . .	6
I.4 Age au premier mariage . . . . .	8
I.5 Migration . . . . .	12
<b>II. Education</b>	<b>15</b>
II.1 Accèsibilité aux écoles . . . . .	15
II.2 Raisons de non Fréquentation . . . . .	17
II.3 Résultat pour l'année scolaire 2019-2020 . . . . .	19
II.4 Raison d'abandon des classe . . . . .	20
II.5 Principales difficultés rencontrées dans le cadre de la scolarité . . . . .	21
<b>III. Caractéristiques sociodémographiques et Education</b>	<b>30</b>
III.1 Niveau d'alphabétisation suivant le sexe . . . . .	30
III.2 Accès à l'école suivant le sexe . . . . .	32
III.3 Niveau d'éducation suivant le sexe . . . . .	33
<b>IV. Création d'indicateur : Indicateur de maîtrise d'une langue (IML)</b>	<b>34</b>
IV.1 Etape 1 : incice simple . . . . .	34
IV.2 Etape 2 : IML pour chaque groupe de langue . . . . .	35
IV. Etape 3 : Indice Globale de Maîtrise de langue . . . . .	35
IV.4 Interprétation de l'indice . . . . .	35

# Introduction

Dans les pays en développement, les enquêtes auprès des ménages sont la source privilégiée pour produire des indicateurs de suivi des conditions de vie des populations. c'est dans ce contexte que dans l'espace UEMOA (Union Economique et Monétaire Ouest Africaine) pour le suivi et l'évaluation de la pauvreté et des conditions de vie des ménages dans chacun des pays membres, il a été mis en place l'Enquête Harmonisée sur les Conditions de vie des Ménages. Et c'est l'analyse des données recueillies dans le cadre de cette enquête au Sénégal qui fait l'objet de cette étude. La présente partie du travail consiste à porter des analyses descriptives sur un échantillon de 13 817 individus sélectionnés dans la précédente partie. Nos différentes analyses porteront essentiellement sur les caractéristiques sociodémographiques et la situation en matière d'éducation au sein de la population sénégalaise résidente.

```
# Chargement de la table de données
## Appel de la bibliothèque nécessaire pour ouvrir des fichier du format dta
library(haven)
library(dplyr)

##
## Attachement du package : 'dplyr'

## Les objets suivants sont masqués depuis 'package:stats':
##
##      filter, lag

## Les objets suivants sont masqués depuis 'package:base':
##
##      intersect, setdiff, setequal, union

## Définissons le chemin d'accès à la tables de données
path <- "C:\\Users\\HP ProBook\\Documents\\Mes Cours ISEP2\\R\\Projet\\Base de données"

## importation des données
df <- read_dta(paste(path,"Base_projetf.dta",sep="\\")) # tables de donnée apurée
head(df)

## # A tibble: 6 x 36
##   key      Sexe Situation_Matri~ Age_au_premier~ Nationalite Ne_dans_sa_loca~
##   <chr>   <dbl+lbl>      <dbl+lbl>      <dbl>      <dbl+lbl>      <dbl+lbl>
## 1 00-01-- 1 [Mas~ 2 [Marié(e) mon~      25 13 [Sénéga~      1 [Oui]
## 2 00-01-- 2 [Fém~ 2 [Marié(e) mon~      15 13 [Sénéga~      1 [Oui]
## 3 00-01-- 1 [Mas~ 1 [Célibataire]      NA 13 [Sénéga~      1 [Oui]
## 4 00-01-- 1 [Mas~ 1 [Célibataire]      NA 13 [Sénéga~      1 [Oui]
## 5 00-01-- 1 [Mas~ 1 [Célibataire]      NA 13 [Sénéga~      1 [Oui]
## 6 00-01-- 1 [Mas~ 1 [Célibataire]      NA 13 [Sénéga~      1 [Oui]
## # ... with 30 more variables: Localite_d_origine_des_migrants <dbl+lbl>,
## #   Principale_cause_de_migration <dbl+lbl>, Annee_de_naissance <dbl>,
## #   Peut_lire_français <dbl+lbl>, Peut_lire_la_langue_locale <dbl+lbl>,
## #   Peut_lire_une_autre_langue <dbl+lbl>, Peut_ecrire_français <dbl>,
## #   Peut_ecrire_la_langue_locale <dbl>, Peut_ecrire_une_autre_langue <dbl>,
## #   comprendre_un_texte_en_français <dbl>,
## #   comprendre_texte_langue_locale <dbl>, ...
```

```
df <- data.frame(df)
```

```
## recodons les variables
```

```
library(labelled)
```

```
variables <- c("Sexe", "Situation_Matrimoniale", "Nationalite"  
              , "Ne_dans_sa_localite_de_residence", "Localite_d_origine_des_migrants", "Principale_cause"  
              , "Peut_lire_la_langue_locale", "Peut_lire_une_autre_langue", "Fait_une_ecole_formelles", "Ra"  
              , "Fait_une_ecole_non_formelle", "Gerant_de_l_ecole" , "Resultat_2019_2020", "Raisons_d_aband"  
              , "Insuffisance_livres_fournitures", "Insuffisance_tablesb_equipements", "Absentisme_enseig"  
              , "Enseignement_pas_satisfaisant", "Effectifs_plethoriques", "Insuffisance_d_enseignants"  
              , "Manque_de_toilettes", "Frequence_des_cotisations", "Salle_de_classe_en_mauvais_etat" )  
df[variables] <- to_factor(df[variables])
```

## I. Caractéristiques sociodémographiques

Dans cette section, on étudiera les caractéristiques sociodémographiques de notre population d'étude à travers des attributs comme la situation matrimoniale, l'âge, l'âge au premier mariage, l'origine et la principale cause de migration. Le sexe et la nationalité des personnes nous aideront pour des analyses par strate.

### I.1 Sexe

```
# fréquences
```

```
knitr::kable(prop.table(table(df["Sexe"])), align = "c")
```

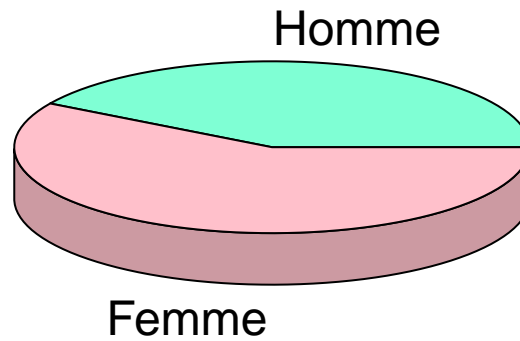
Sexe	Freq
Masculin	0.414996
Féminin	0.585004

```
# graphique en secteur
```

```
library(plotrix)
```

```
pie3D(table(df["Sexe"]), labels = c("Homme", "Femme") , col = c("aquamarine", "pink"), main = "Sexe")
```

## Sexe



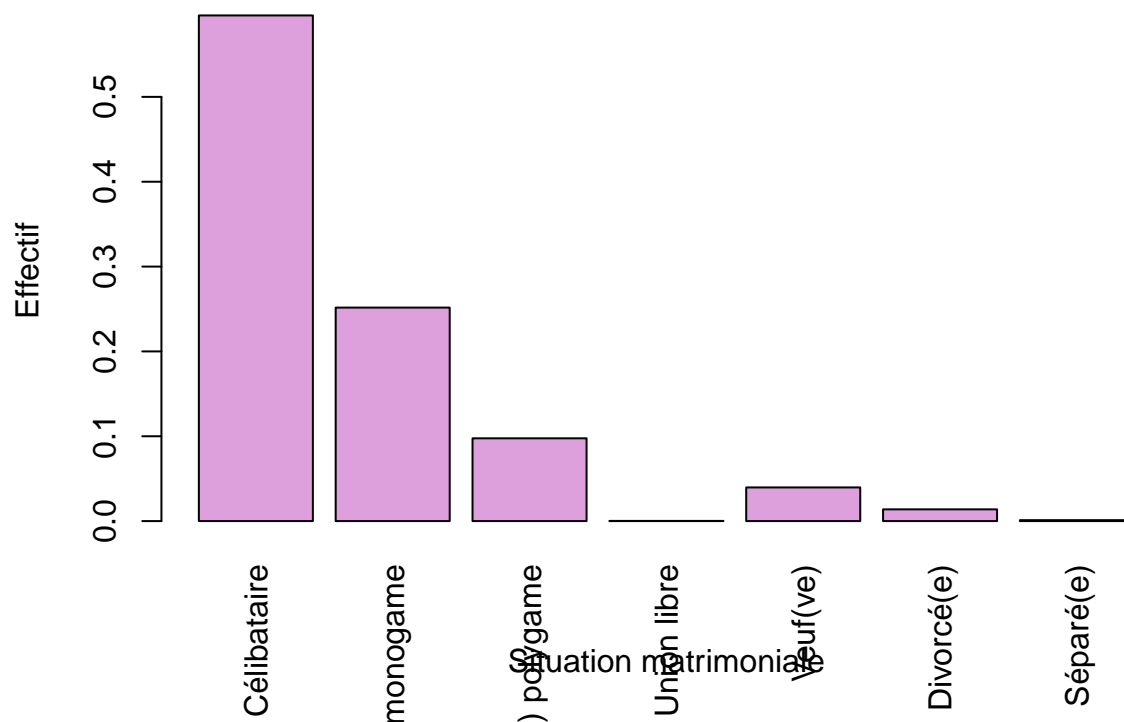
L'analyse du graphique ci-dessus confirme la tendance démographique de la sous région avec une plus grande proportion de femmes. En effet, les femmes occupent 58,5 % de notre population d'étude.

## I.2 Situation matrimoniale

```
## situation Matrimoniale
### Fréquences
knitr::kable(prop.table(table(df["Situation_Matrimoniale"])), align = "c")
```

Situation_Matrimoniale	Freq
Célibataire	0.5960049
Marié(e) monogame	0.2516465
Marié(e) polygame	0.0975610
Union libre	0.0002171
Veuf(ve)	0.0396613
Divorcé(e)	0.0138236
Séparé(e)	0.0010856

```
### diagramme en baton
barplot(prop.table(table(df["Situation_Matrimoniale"])), xlab="Situation matrimoniale", ylab="Effectif", )
```



Le graphique ci-dessus montre que sur notre échantillon plus de la moitié des individus (59,53 %) sont célibataires et qu'il y a très peu de personnes en union libres comme séparés. Quant aux mariés, ils sont plus monogames que polygames. Aussi, constate-t-on que les veufs et les divorcés sont moins présents que les célibataires et les mariés mais ils sont plus nombreux que les divorcés et les séparés.

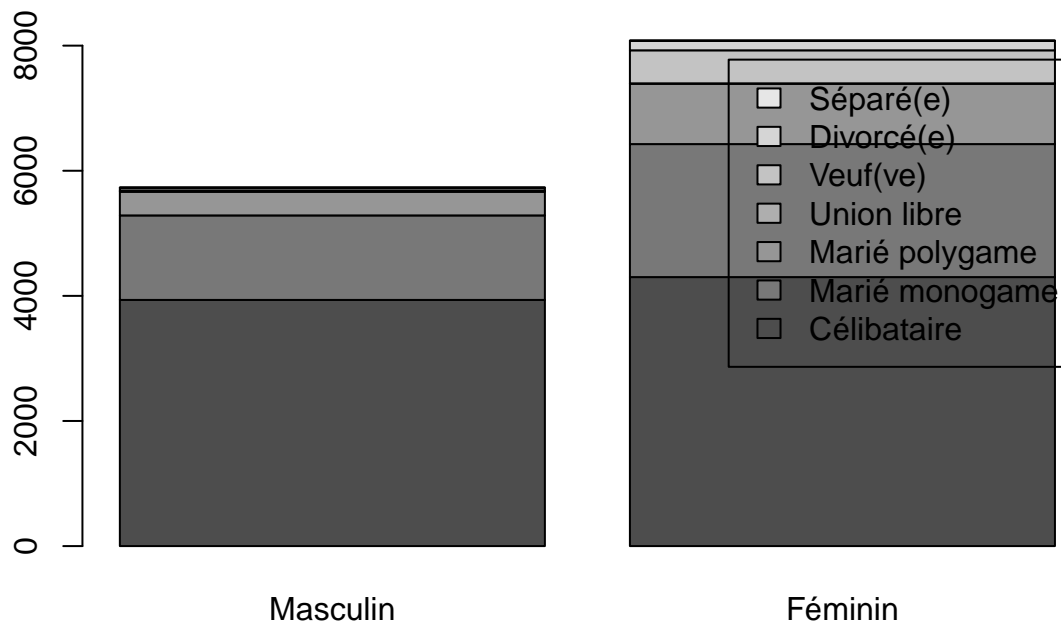
```
#Regardons la variable situations matrimoniale en relation avec le sexe
tab2 <- table(df[, "Situation_Matrimoniale"], df[, "Sexe"])
knitr::kable(tab2, align = "c")
```

	Masculin	Féminin
Célibataire	3935	4300
Marié(e) monogame	1351	2126
Marié(e) polygame	380	968
Union libre	0	3
Veuf(ve)	20	528
Divorcé(e)	40	151
Séparé(e)	8	7

```
#Représentation graphique
```

```
barplot(tab2, main = "Situation matrimoniale en fonction du sexe", legend.text = c("Célibataire", "Marié m
```

## Situation matrimoniale en fonction du sexe



```
chisq.test(tab2) # TEST DE KHI2
```

```
##
## Pearson's Chi-squared test
##
## data:  tab2
## X-squared = 601.95, df = 6, p-value < 2.2e-16
```

Pour les femmes de même que les hommes, les célibataires et les mariés dominent en nombre alors que les séparés et les personnes en unions libres sont moins nombreux. Cependant force est de remarquer qu'il y a plus de veuves que de veuf (96,35% des veufs sont des veuves). Tandance qui semble se perpétuer parmi les divorcés nous poussant à suspecter un lien entre la situation matrimoniale et le sexe. Suspension qui est confirmée par un test d'indépendance de Khi2 ( $p\text{-value} < 2.2e-16 < 0.05 = \text{seuil de significativité}$ )

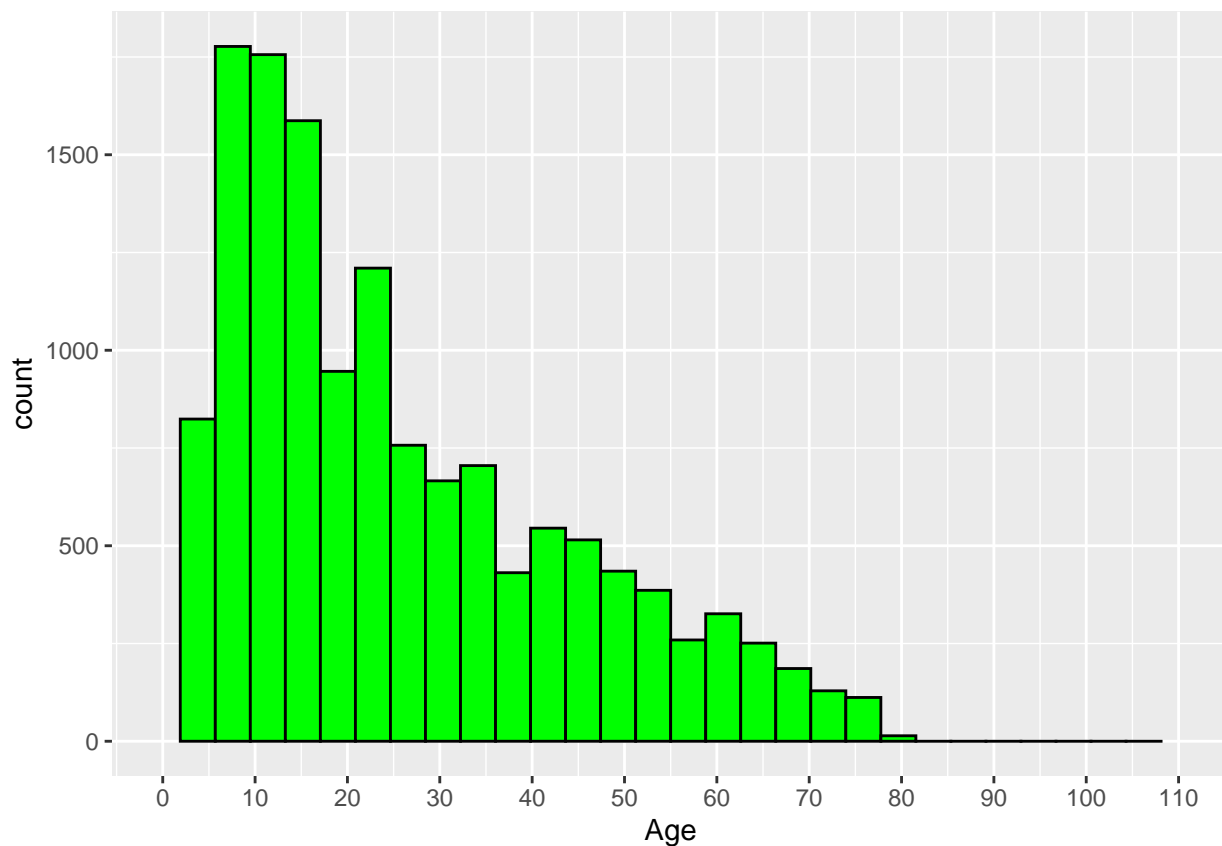
### I.3 Age

```
## Age
### statistiques descriptives
knitr::kable(summary(df["Age"]), align = "c")
```

Age
Min. : 4.00
1st Qu.:11.00
Median :21.00
Mean :25.93
3rd Qu.:37.00
Max. :78.00

```
### histogramme
library(ggplot2)
#### histogramme classique
ggplot(df, aes(x=Age)) +
  geom_histogram(color="black", fill="green") +
  theme(legend.position="top")+
  scale_x_continuous(breaks=seq(0,110,10), limits=c(0,110))
```

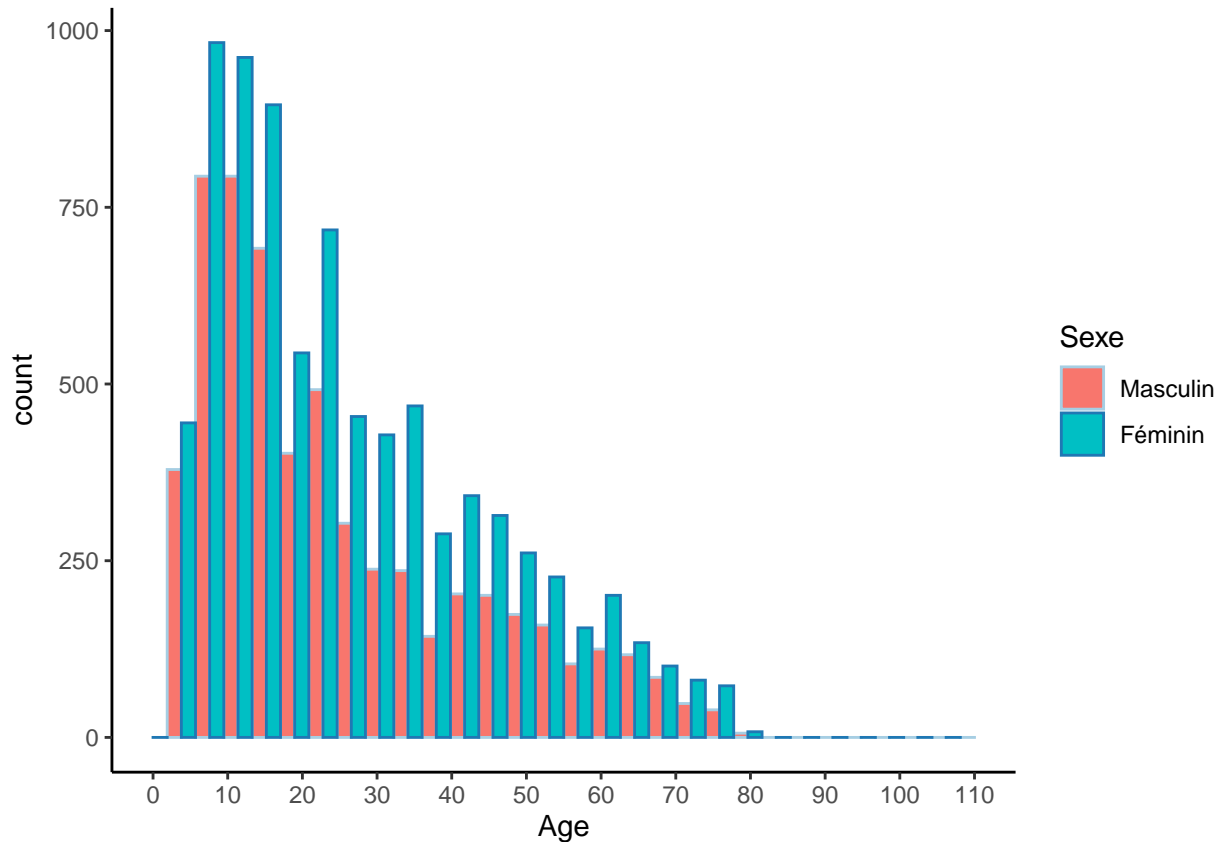
## 'stat\_bin()' using 'bins = 30'. Pick better value with 'binwidth'.



```
#### Histogramme par sexe
ggplot(df, aes(x=Age, color=Sexe, fill=Sexe)) +
  geom_histogram(position="dodge") +
  theme(legend.position="top")+
  scale_color_brewer(palette="Paired") +
```

```
theme_classic()+
scale_x_continuous(breaks=seq(0,110,10), limits=c(0,110))
```

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



Notre population est jeune et composée de personnes âgées entre 4 et 78 ans. Plus de la moitié de cette population soit les trois quart pour être précis ont moins de 37 ans. Le second graphique nous montre que les femmes vivent généralement plus longtemps que les hommes.

## I.4 Age au premier mariage

```
## Age au premier mariage
### statistiques descriptives
knitr::kable(summary(df["Age_au_premier_mariage"]), align = "c")
```

Age_au_premier_mariage
Min. :12.00
1st Qu.:18.00
Median :20.00
Mean :21.85
3rd Qu.:25.00



Age_au_premier_mariage
Max. :60.00
NA's :8307

```
## description de age moyen de mariage par sexe
fem <- df[which(df["Sexe"] == "Féminin"),]
hom <- df[which(df["Sexe"] == "Masculin"),]
print("Age au premier mariage chez les femmes")
```

```
## [1] "Age au premier mariage chez les femmes"
```

```
knitr::kable(summary(fem["Age_au_premier_mariage"]), align = "c")
```

Age_au_premier_mariage
Min. :12.00
1st Qu.:17.00
Median :19.00
Mean :20.07
3rd Qu.:22.00
Max. :45.00
NA's :4354

```
print("Age au premier mariage chez les hommes")
```

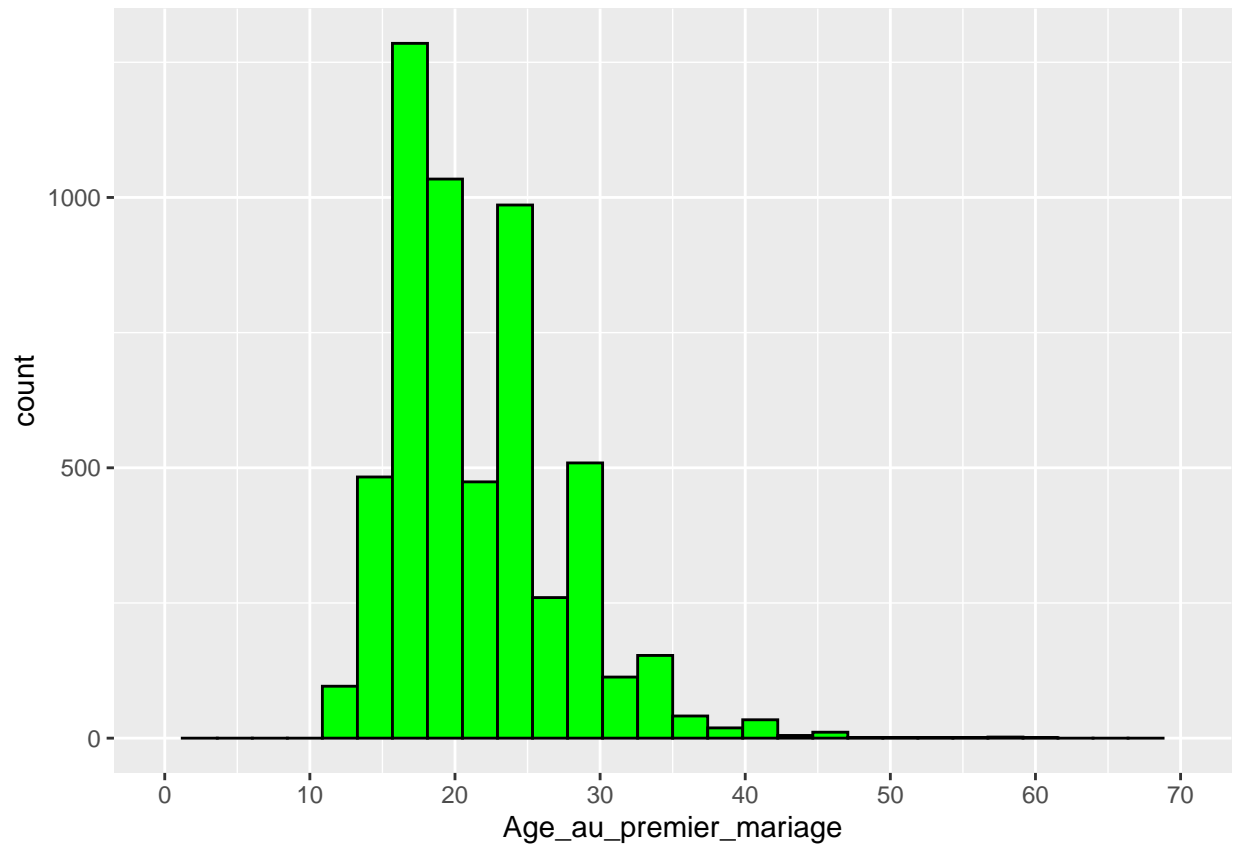
```
## [1] "Age au premier mariage chez les hommes"
```

```
knitr::kable(summary(hom["Age_au_premier_mariage"]), align = "c")
```

Age_au_premier_mariage
Min. :13.00
1st Qu.:22.00
Median :25.00
Mean :25.59
3rd Qu.:29.00
Max. :60.00
NA's :3953

```
### histogramme
library(ggplot2)
#### histogramme classique
ggplot(df, aes(x=Age_au_premier_mariage)) +
  geom_histogram(color="black", fill="green") +
  theme(legend.position="top")+
  scale_x_continuous(breaks=seq(0,70,10), limits=c(0,70))
```

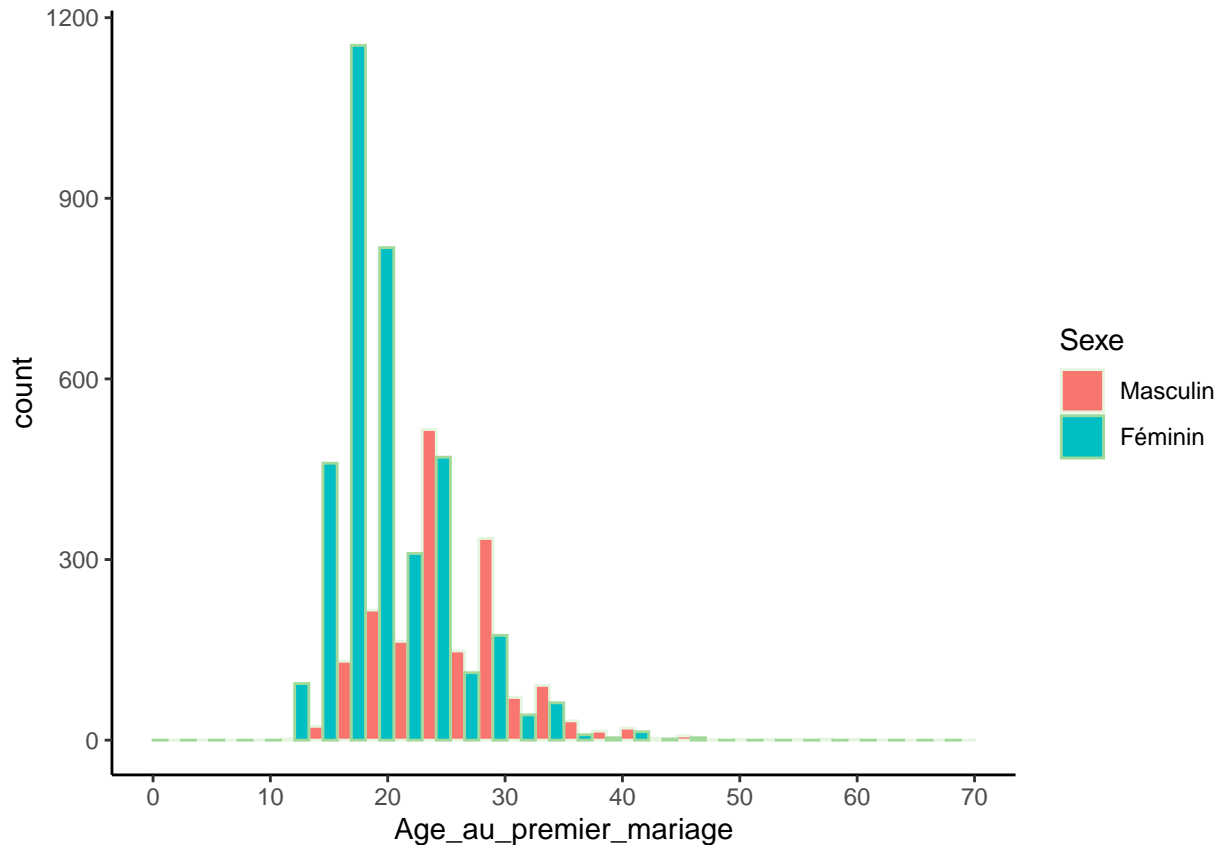
```
## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
```



#### #### Histogramme par sexe

```
ggplot(df, aes(x=Age au premier mariage, color=Sexe, fill=Sexe)) +
  geom_histogram(position="dodge") +
  theme(legend.position="top")+
  scale_color_brewer(palette="paired") +
  theme_classic()+
  scale_x_continuous(breaks=seq(0,70,10), limits=c(0,70))
```

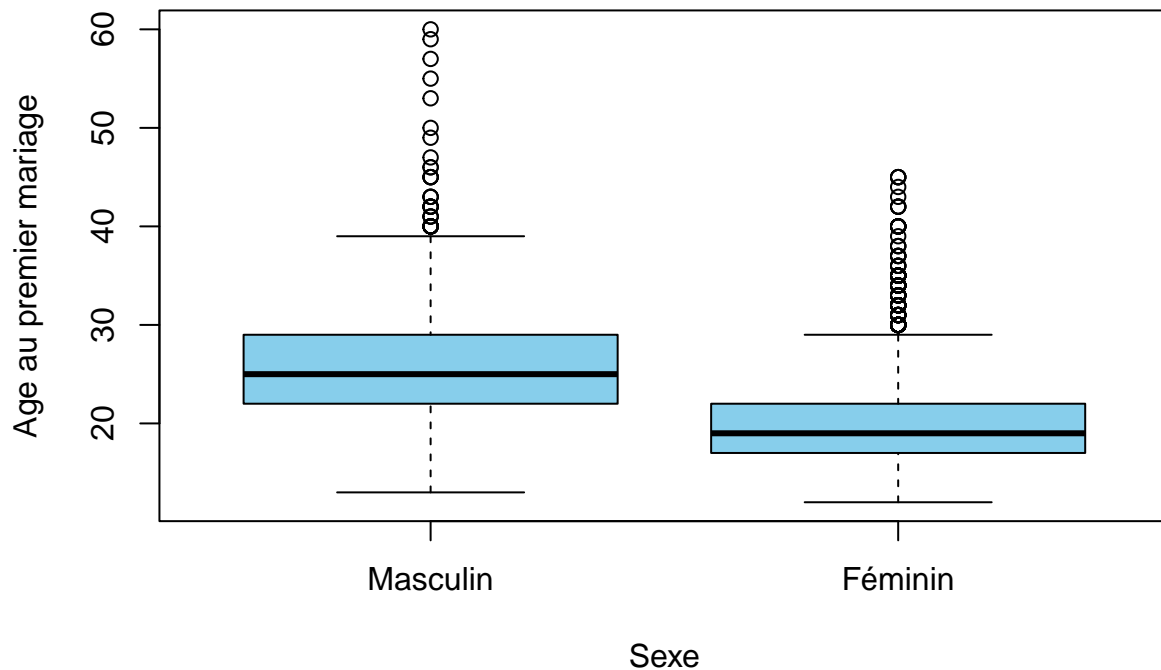
## 'stat\_bin()' using 'bins = 30'. Pick better value with 'binwidth'.



Il ressort de l'analyse des graphiques et des chiffres ci-dessus que les gens se marient (pour la 1ère fois) en moyenne à 22 ans. Cet âge moyen est plus bas de 2 ans chez les femmes (20 ans) alors qu'il est de 26 ans pour les hommes soit 4 ans de plus. Cela nous permet avec l'appui du graphique ci-après et des résultats du test de student ( $p\text{-value} < 2.2e-16 < 0.05 = \text{seuil de significativité}$ ) d'affirmer que les femmes se marient généralement plus tôt que les hommes. Cependant force est de remarquer que près d'un quart des personnes qui sont mariés, l'ont fait avant leur 18 ans âge légal du mariage pour les filles comme les garçons selon la Convention internationale des droits de l'enfant. Par ailleurs, on remarque que la situation est plus alarmante chez les femmes où 25% se sont marié avant 17 ans. En se référant à l'âge légal du mariage au Sénégal qui est de 16 ans pour les femmes et 18 ans chez les hommes puis aux normes internationales, on peut dire que le Sénégal est encore à la traîne en matière d'éradication des mariages précoces en général et chez la femme surtout.

```
boxplot(df$Age au premier mariage ~ df$Sexe,col="skyblue",
        main="Age au premier mariage selon l'Age",
        ylab="Age au premier mariage",
        xlab="Sexe" )
```

## Age au premier mariage selon l'Age



```
# Faisons un test de student pour comparer l'âge moyen de mariage chez les filles et les garçons afin d
t.test(df$Age_au_premier_mariage ~ df$Sexe)
```

```
##
## Welch Two Sample t-test
##
## data: df$Age_au_premier_mariage by df$Sexe
## t = 35.154, df = 3056.1, p-value < 2.2e-16
## alternative hypothesis: true difference in means between group Masculin and group Féminin is not equa
## 95 percent confidence interval:
##  5.207260 5.822449
## sample estimates:
## mean in group Masculin mean in group Féminin
##          25.58619          20.07133
```

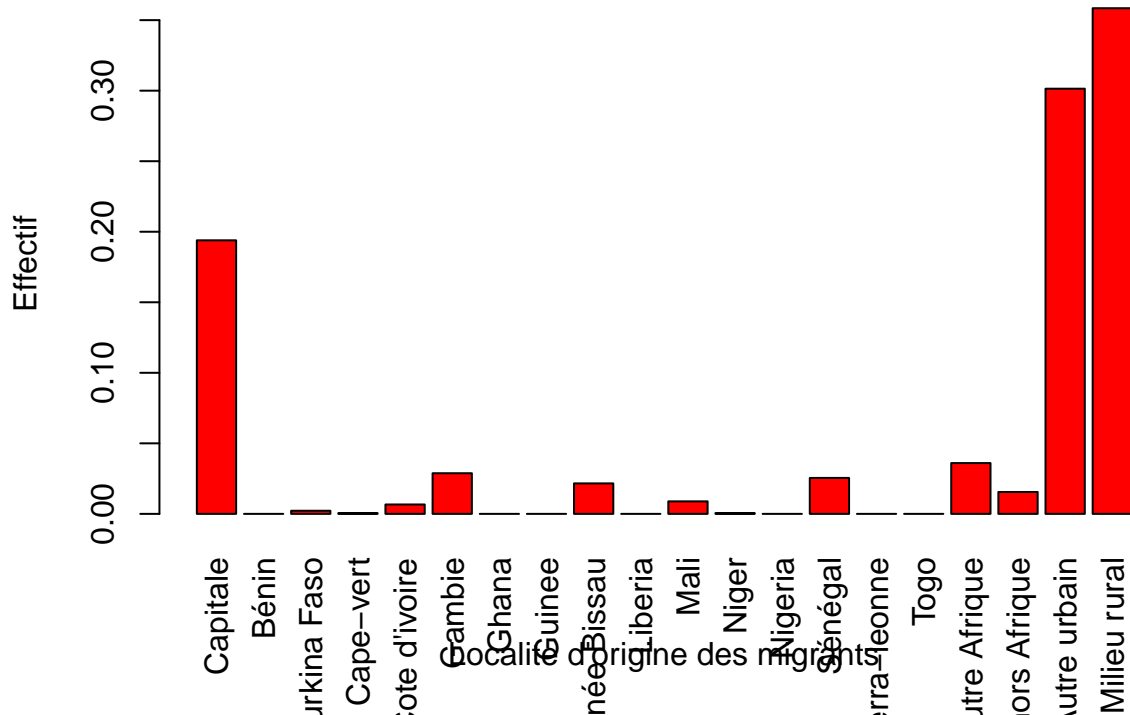
## I.5 Migration

### I.5.1 Origine des migrants

```
## origine des migrants
### Fréquences
prop.table(table(df["Localite_d_origine_des_migrants"]))
```

```
## Localite_d_origine_des_migrants
##          Capitale          Bénin          Burkina Faso
##          0.1939058172          0.0000000000          0.0022160665
##          Cape-vert          Cote d'Ivoire          Gambie
##          0.0005540166          0.0066481994          0.0288088643
##          Ghana          Guinee          Guinée Bissau
##          0.0000000000          0.0000000000          0.0216066482
##          Liberia          Mali          Niger
##          0.0000000000          0.0088642659          0.0005540166
##          Nigeria          Sénégal          Serra-leonne
##          0.0000000000          0.0254847645          0.0000000000
##          Togo          Autre Afrique Autre pays hors Afrique
##          0.0000000000          0.0360110803          0.0155124654
##          Autre urbain          Milieu rural
##          0.3013850416          0.3584487535
```

```
### diagramme en baton
barplot(prop.table(table(df["Localite_d_origine_des_migrants"])),
        ,xlab="Localite d'origine des migrants",ylab="Effectif",col = "red", las = 3)
```



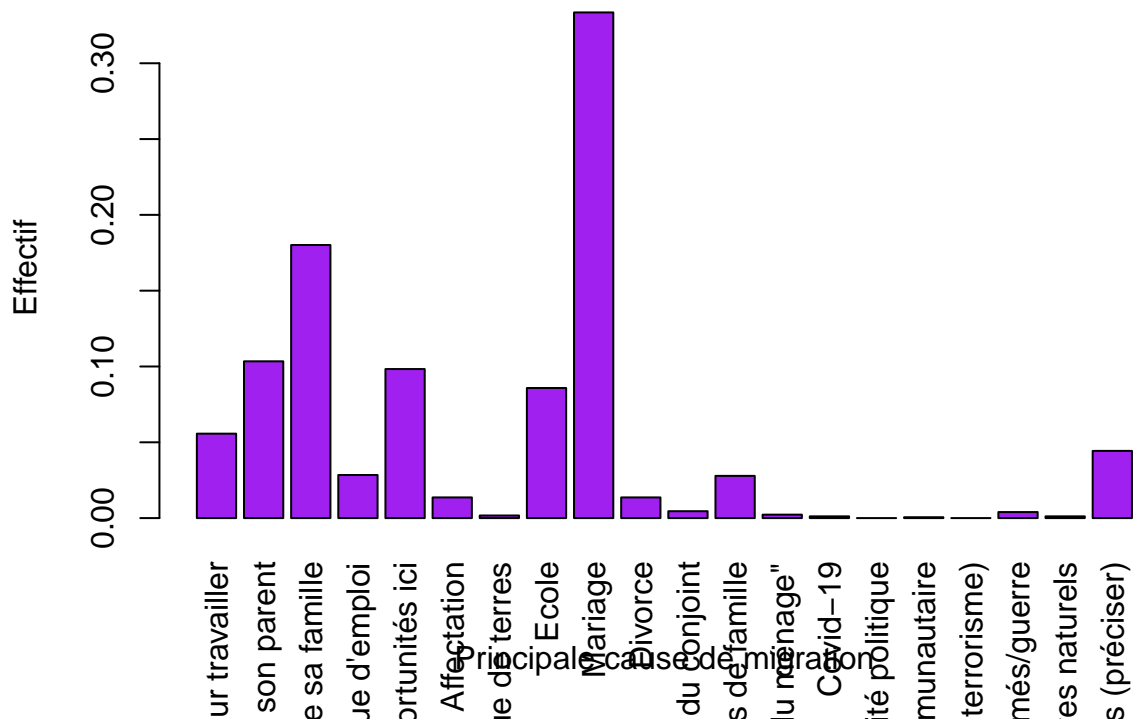
Dans notre population, pour une localité donnée, les principaux émigrés proviennent du milieu rural, de la capitale et de la zone urbaine. On y trouve des non-sénégalais mais en très faible nombre. Parmi ces derniers, les Gambiens, les bissau guinéens, les ivoiriens et les personnes provenant d'autres continents sont les plus représentés.

## I.5.2 Principales cause de migration

```
## Principale cause de migration
### Fréquences
prop.table(table(df["Principale_cause_de_migration"]))
```

```
## Principale_cause_de_migration
## Envoyé par sa famille pour travailler      Est venu avec son parent
##              0.0556818182                  0.1034090909
##      Suivre/rejoindre sa famille            Manque d'emploi
##              0.1801136364                  0.0284090909
##      Meilleures Opportunités ici            Affectation
##              0.0982954545                  0.0136363636
##      Manque de terres                       Ecole
##              0.0017045455                  0.0857954545
##      Mariage                               Divorce
##              0.3335227273                  0.0136363636
##      Perte du conjoint                     Autres problèmes de famille
##              0.0045454545                  0.0278409091
##      Maladie d'un membre du ménage"        Covid-19
##              0.0022727273                  0.0011363636
##      Instabilité politique                 Conflit intercommunautaire
##              0.0000000000                  0.0005681818
##      Insécurité (banditisme, terrorisme)    Conflits armés/guerre
##              0.0000000000                  0.0039772727
##      Désastres naturels                     Autres (préciser)
##              0.0011363636                  0.0443181818
```

```
### diagramme en baton
barplot(prop.table(table(df["Principale_cause_de_migration"])),
        ,xlab="Principale cause de migration",ylab="Effectif",col = "purple", las = 3)
```



La liste des principales causes de migration est longue mais la principale est le mariage. Mis à part cela, la plupart des personnes se déplacent pour chercher du travail, accompagner leur parent, rejoindre leur famille, étudier ou chercher de meilleures opportunités.

## II. Education

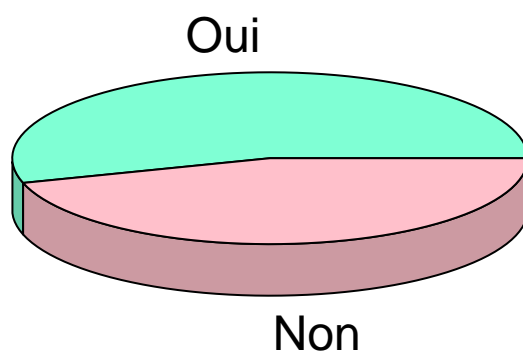
Le but de cette partie, est évaluer le secteur éducatif de notre population.

### II.1 Accèsibilité aux écoles

```
## Accèsibilité aux écoles
acces <- c("Fait_une_ecole_formelles", "Fait_une_ecole_non_formelle")
for (x in acces) {
  ### Fréquences
  print(prop.table(table(df[which(df$Age < 16 & df$Age >= 6 ), x])))
  pie3D(table(df[which(df$Age < 16 & df$Age >= 6 ), x]), labels = c("Oui", "Non"), col = c("aquamarine", "p
}

##
##      Oui      Non
## 0.5464418 0.4535582
```

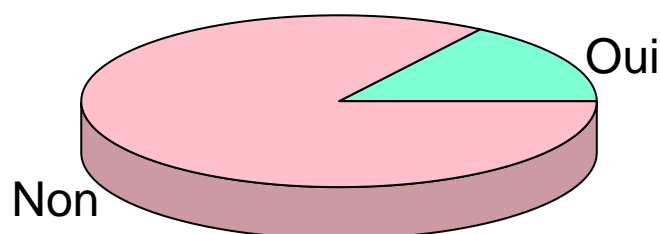
## Fait\_une\_ecole\_formelles



```
##  
##      Oui      Non  
## 0.1578096 0.8421904
```



## Fait\_une\_ecole\_non\_formelle



On peut lire dans l'article premier de la Loi 2004-37 du 15 décembre 2004 modifiant et complétant la Loi d'orientation de l'Éducation nationale no 91-22 du 16 février 1991 du Sénégal ce qui suit : "Il est fait obligation aux parents, dont les enfants atteignent l'âge de 6 ans, de les inscrire dans une école publique ou privée. Les parents sont tenus de s'assurer de l'assiduité de leurs enfants jusqu'à l'âge de 16 ans." En référence à cette loi et avec les résultats présentés par les graphiques ci-dessus, nous pouvons dire que la situation dans notre population d'étude est critique car encore 45% des enfants de cette tranche d'âge n'ont pas accès à une formation formelle. Mais qu'est-ce qui explique cela ?

## II.2 Raisons de non Fréquentation

### ### Fréquences

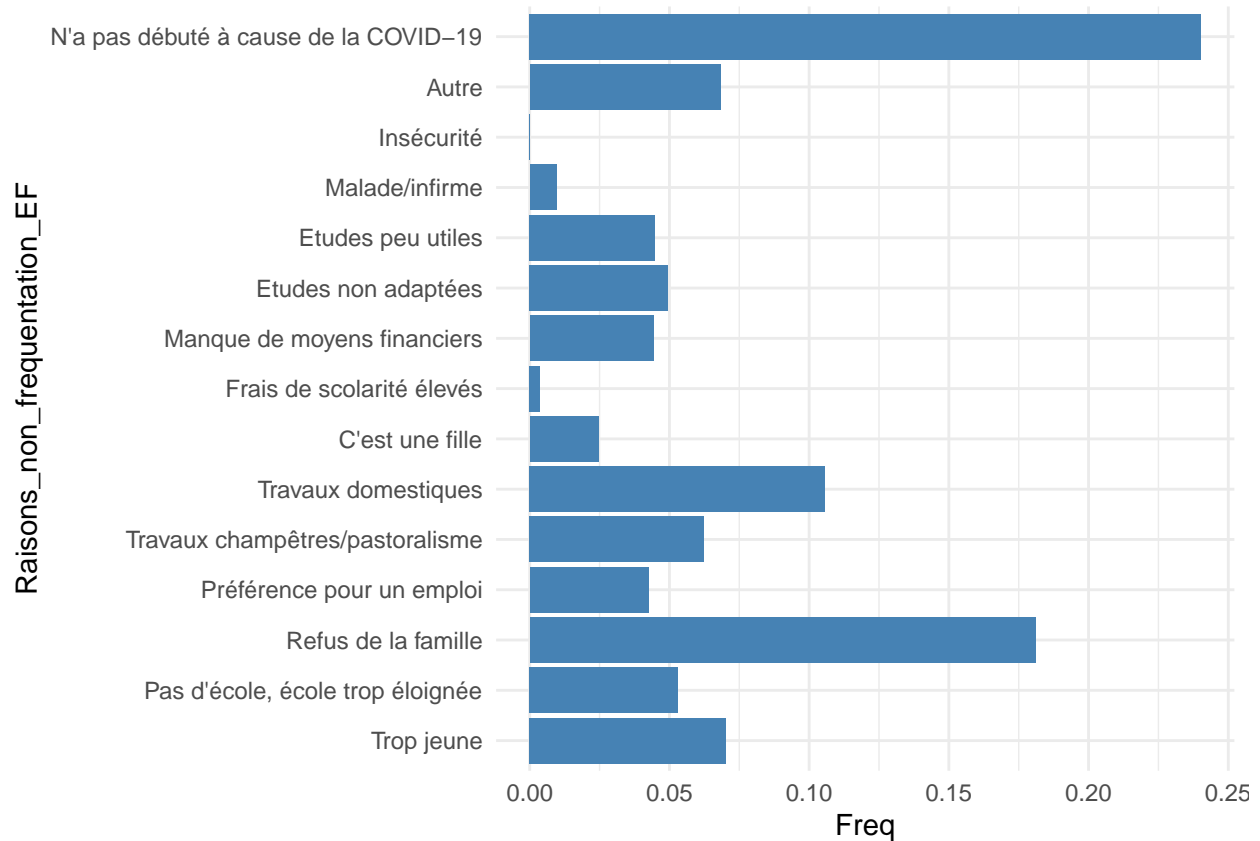
```
y1 <- data.frame(prop.table(table(df["Raisons_non_frequentation_EF"])))  
print(y1)
```

##	Raisons_non_frequentation_EF	Freq
## 1	Trop jeune	0.0702749353
## 2	Pas d'école, école trop éloignée	0.0530144249
## 3	Refus de la famille	0.1809887807
## 4	Préférence pour un emploi	0.0425348292
## 5	Travaux champêtres/pastoralisme	0.0623844162
## 6	Travaux domestiques	0.1056589816
## 7	C'est une fille	0.0246578720
## 8	Frais de scolarité élevés	0.0036986808
## 9	Manque de moyens financiers	0.0442608803

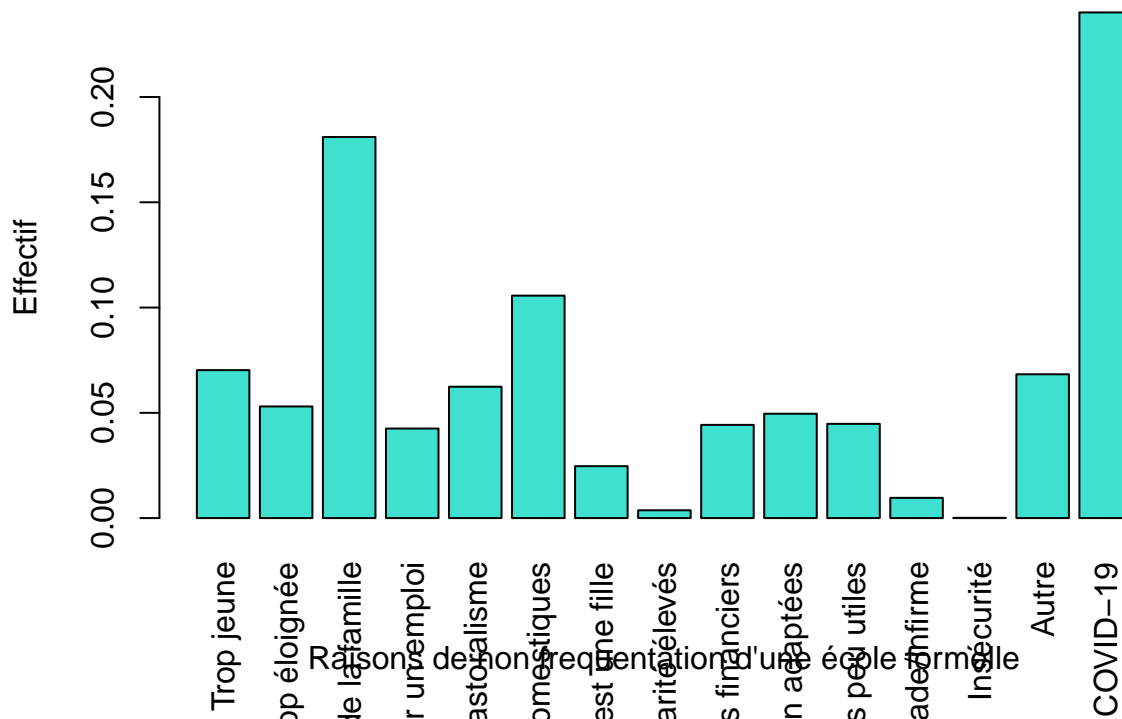
```
## 10          Etudes non adaptées 0.0495623228
## 11          Etudes peu utiles 0.0447540377
## 12          Malade/infirmes 0.0096165701
## 13          Insécurité 0.0001232894
## 14          Autre 0.0683023055
## 15 N'a pas débuté à cause de la COVID-19 0.2401676735
```

```
### diagramme en baton
```

```
ggplot(data=y1, aes(x=Freq, y=Raisons_non_frequentation_EF)) +
  geom_bar(stat="identity", fill="steelblue")+ theme_minimal()
```



```
barplot(prop.table(table(df["Raisons_non_frequentation_EF"])))
, xlab="Raisons de non frequentation d'une école formelle", ylab="Effectif", col = "turquoise", las=1)
```



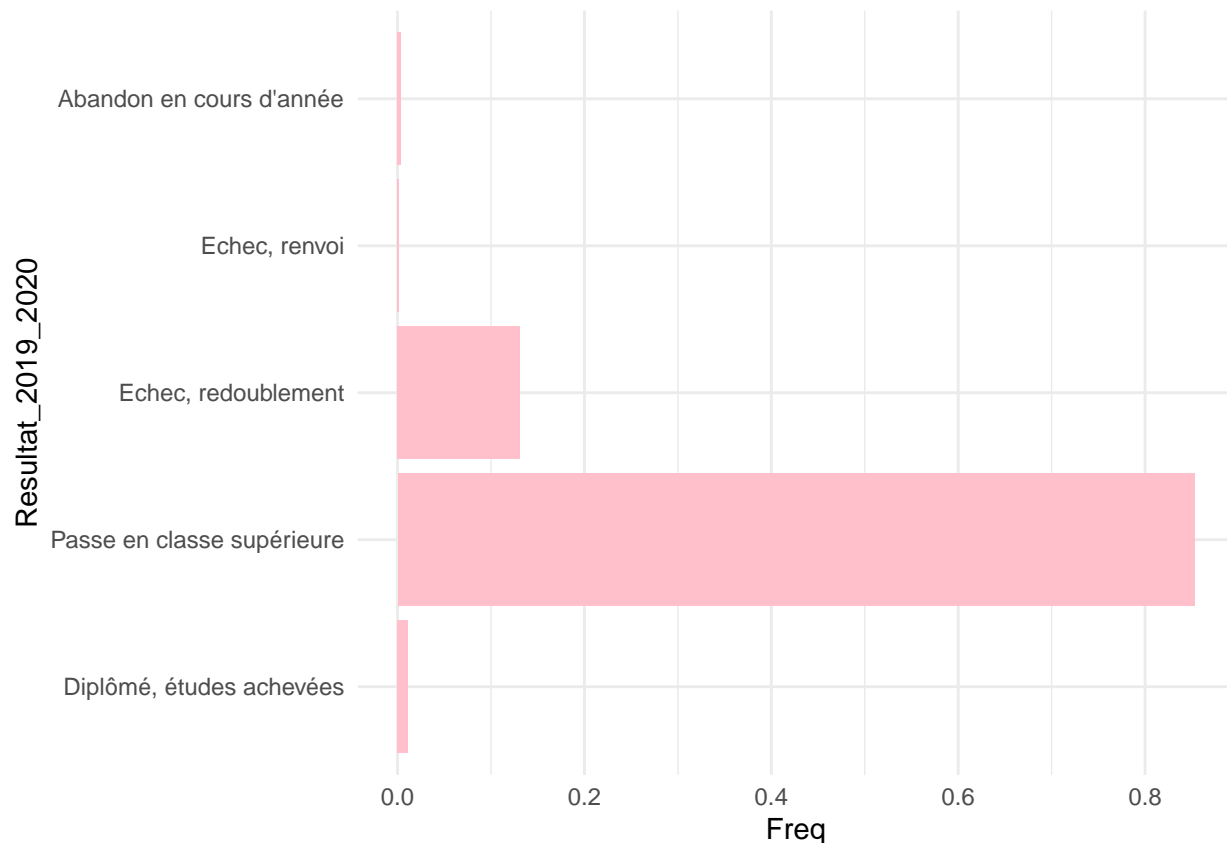
Le graphique ci-dessus tente d'apporter une réponse à la précédente question. On peut y voir que pour le compte de l'année scolaire 2019-2020, La principale raison de non fréquentation de l'école était la COVID-19, ce qui est assez intuitif vu que le monde traversait une crise sanitaire sans précédent. Par contre on relève d'autres raisons entre autres : le refus de la famille, les travaux domestiques, le manque d'établissements ou l'éloignement de ces derniers etc.

### II.3 Résultat pour l'année scolaire 2019-2020

```
y1 <- data.frame(prop.table(table(df["Resultat_2019_2020"])))
print(y1)
```

```
##          Resultat_2019_2020          Freq
## 1  Diplômé, études achevées 0.011075720
## 2  Passe en classe supérieure 0.852975243
## 3      Echec, redoublement 0.131098885
## 4      Echec, renvoi 0.001158245
## 5  Abandon en cours d'année 0.003691907
```

```
### diagramme en baton
ggplot(data=y1, aes(x=Freq, y=Resultat_2019_2020)) +
  geom_bar(stat="identity", fill="pink") + theme_minimal()
```



Certes, il y a des redoublements, des abandons, des renvois mais le taux des élèves /étudiants qui ont passé en classe supérieure est très élevé. L'enseignement souffre d'une part des manquements nonobstant que les résultats de l'année scolaire 2019-2020 sont très satisfaisants.

## II.4 Raison d'abandon des classe

Bien qu'il y est eu peu d'abandon au cours de l'année scolaire 2019-2020, il reste légitime de se questionner sur les causes de ces abandons. Ainsi, il ressort que La grossesse, la difficulté des études, le manque de moyens financiers, les travaux domestiques et les travaux champêtres sont des raisons qui poussent à l'abandon des classes. Par ailleurs, on note qu'aucune fille n'a abandonné les classes sous prétexte qu'elle était une fille, ce qui témoigne d'une avancée en matière d'émancipation de la femme et de lutte contre le sexisme.

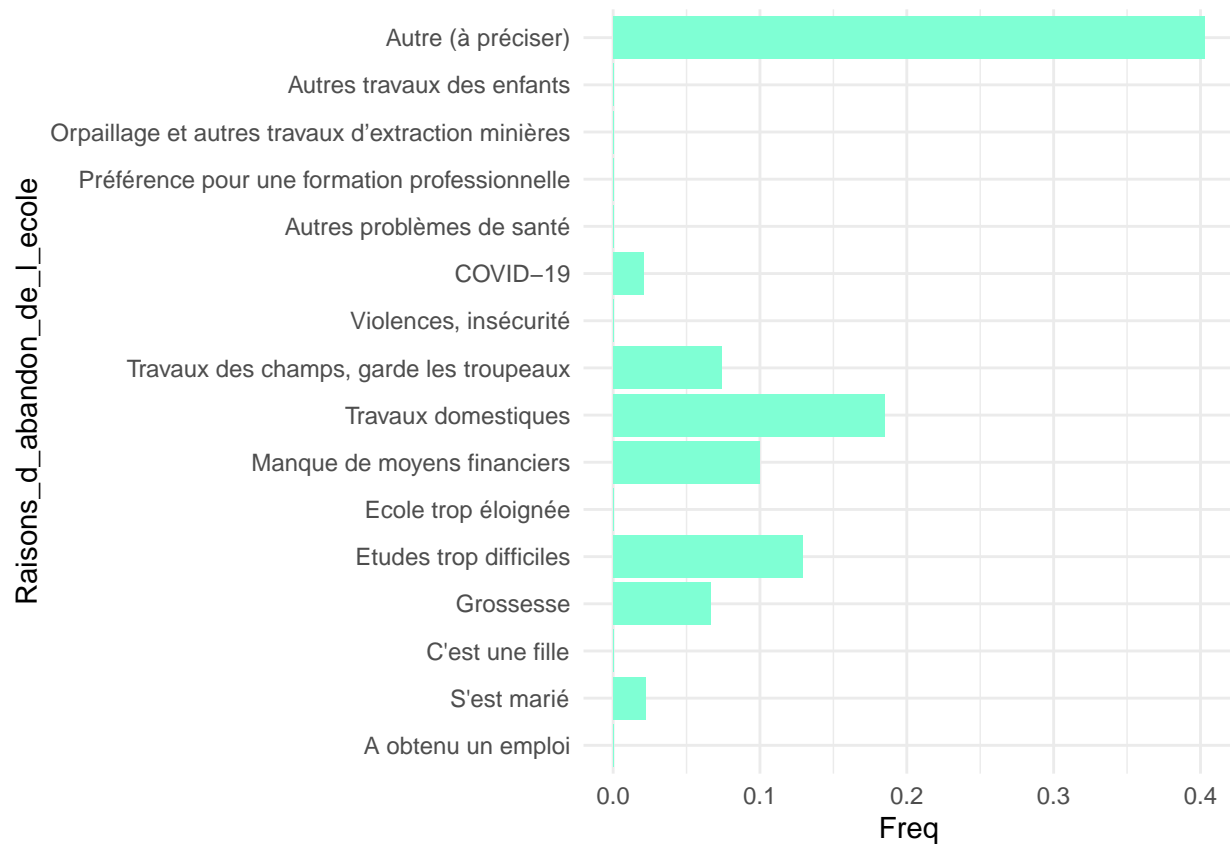
```
y1 <- data.frame(prop.table(table(df["Raisons_d_abandon_de_l_ecole"])))
print(y1)
```

```
##           Raisons_d_abandon_de_l_ecole      Freq
## 1                A obtenu un emploi 0.00000000
## 2                  S'est marié 0.02214443
## 3                C'est une fille 0.00000000
## 4                  Grossesse 0.06628110
## 5          Etudes trop difficiles 0.12890952
## 6          Ecole trop éloignée 0.00000000
## 7          Manque de moyens financiers 0.09968800
## 8            Travaux domestiques 0.18506963
## 9    Travaux des champs, garde les troupeaux 0.07411917
```

```
## 10                               Violences, insécurité 0.00000000
## 11                               COVID-19 0.02085077
## 12                               Autres problèmes de santé 0.00000000
## 13      Préférence pour une formation professionnelle 0.00000000
## 14 Orpaillage et autres travaux d'extraction minières 0.00000000
## 15                               Autres travaux des enfants 0.00000000
## 16                               Autre (à préciser) 0.40293737
```

```
### diagramme en baton
```

```
ggplot(data=y1, aes(x=Freq, y=Raisons_d_abandon_de_l_ecole)) +
  geom_bar(stat="identity", fill="aquamarine")+ theme_minimal()
```



## II.5 Principales difficultés rencontrées dans le cadre de la scolarité

Il est important de noter que dans le cadre scolaire, on fait face à des problèmes tels que l'insuffisance des fournitures scolaires, des problèmes au niveau des équipements scolaires et des problèmes au niveau des enseignants. Cependant, on remarque que hormis l'insuffisance des livres et fournitures scolaire, plus de 72% des personnes ayant été à l'école ne rencontrent pas les autres difficultés.

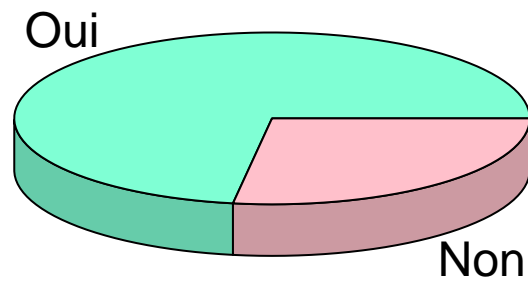
```
## Principales difficultés rencontrées dans le cadre de la scolarité
```

```
difficulte <- c("Insuffisance_livres_fournitures", "Insuffisance_tablesb_equipements", "Absentisme_ensei",
  "Enseignement_pas_satisfaisant", "Effectifs_plethoriques", "Insuffisance_d_enseignants",
  "Manque_de_toilettes", "Frequence_des_cotisations", "Salle_de_classe_en_mauvais_etat")
```

```
for(x in difficulte){
  print(prop.table(table(df[x])))
  pie3D(table(df[x]),labels = c("Oui","Non") ,col = c("aquamarine","pink"), main = x)
}
```

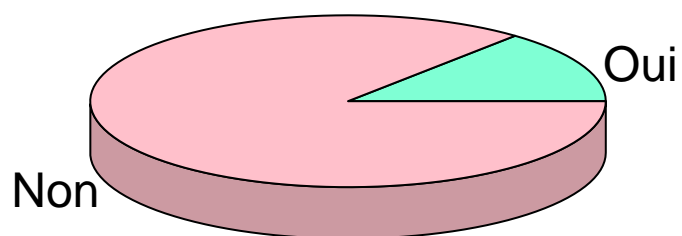
```
## Insuffisance_livres_fornitures
##      Oui      Non
## 0.7258104 0.2741896
```

### Insuffisance\_livres\_fornitures



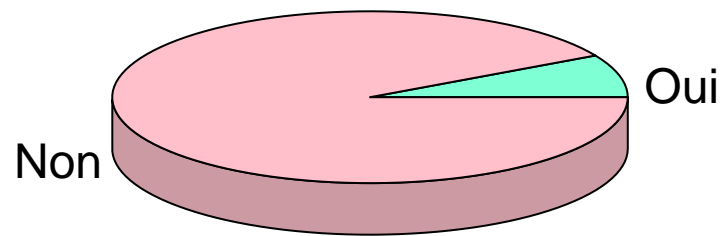
```
## Insuffisance_tablesb_equipements
##      Oui      Non
## 0.1377701 0.8622299
```

## Insuffisance\_tablesb\_equipements



```
## Absenteisme_enseignants_greve
##      Oui      Non
## 0.08067289 0.91932711
```

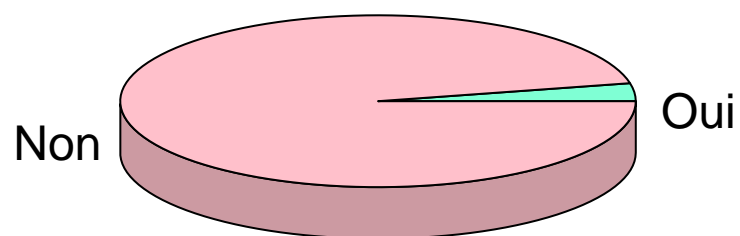
## Absenteisme\_enseignants\_greve



```
## Enseignement_pas_satisfaisant
##      Oui      Non
## 0.0336444 0.9663556
```

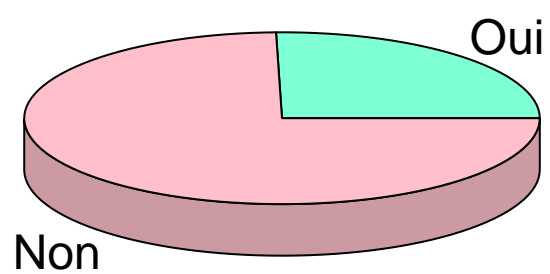


## Enseignement\_pas\_satisfaisant



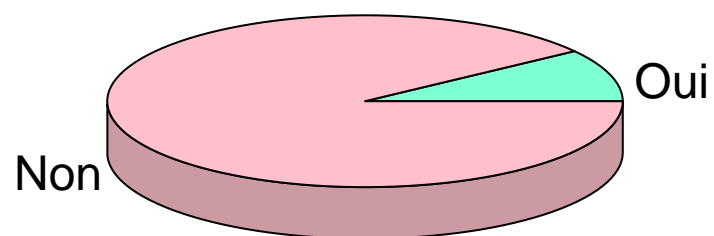
```
## Effectifs_plethoriques
##      Oui      Non
## 0.2536837 0.7463163
```

## Effectifs\_plethoriques



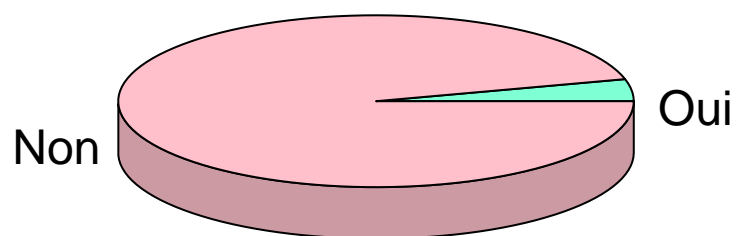
```
## Insuffisance_d_enseignants
##      Oui      Non
## 0.09896857 0.90103143
```

## Insuffisance\_d\_enseignants



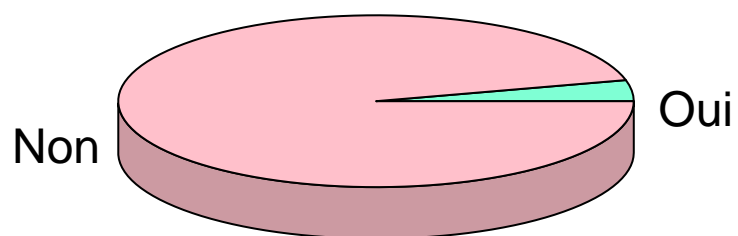
```
## Manque_de_toilettes
##      Oui      Non
## 0.04138016 0.95861984
```

## Manque\_de\_toilettes



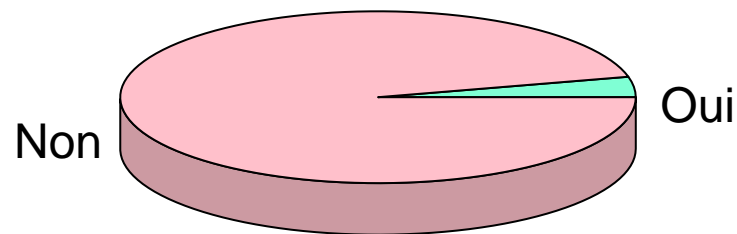
```
## Frequence_des_cotisations
##      Oui      Non
## 0.03953831 0.96046169
```

## Frequence\_des\_cotisations



```
## Salle_de_classe_en_mauvais_etat
##      Oui      Non
## 0.03831041 0.96168959
```

## Salle\_de\_classe\_en\_mauvais\_etat



### III. Caractéristiques sociodémographiques et Education

La présente section va mettre en rapport certaines caractéristiques sociodémographiques et l'éducation de notre population d'étude.

#### III.1 Niveau d'alphabétisation suivant le sexe

```
# Abordons le niveau d'alphabétisation selon le sexe
```

```
# croisons sexe et peut_lire_français
```

```
tab5 <- table(df$Peut_lire_français,df$Sexe)  
print(tab5)
```

```
##  
##      Masculin Féminin  
##  Oui      2429    3249  
##  Non      3305    4834
```

```
library("questionr")  
print(cprop(tab5))
```

```
##
##      Masculin Féminin Ensemble
## Oui    42.4    40.2    41.1
## Non    57.6    59.8    58.9
## Total 100.0   100.0   100.0
```

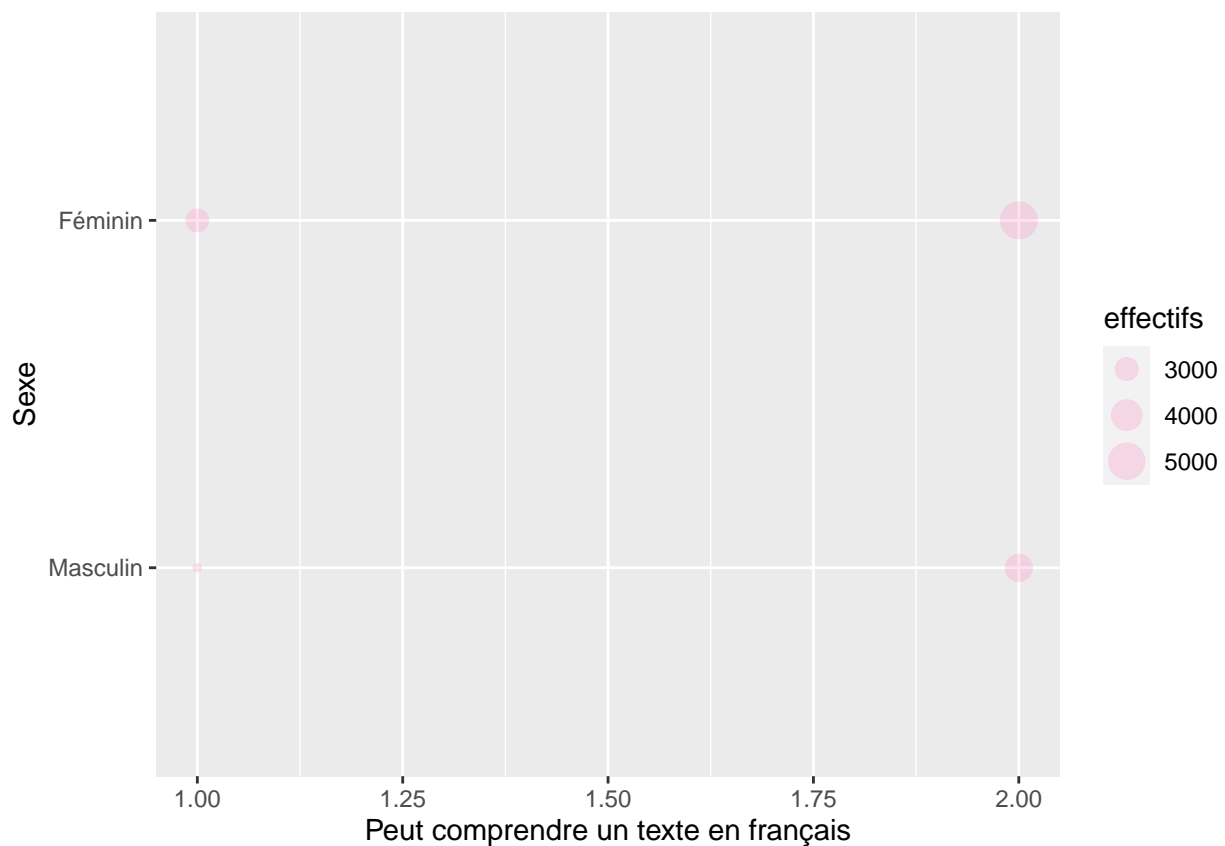
```
# Croisons sexe et peut ?crire en fran?ais
```

```
tab6 <- table(df$Peut_ecrire_français,df$Sexe)
print(cprop(tab6))
```

```
##
##      Masculin Féminin Ensemble
## 1    40.6    38.7    39.5
## 2    59.4    61.3    60.5
## Total 100.0   100.0   100.0
```

```
# croisons sexe et comprendre un texte en fran?ais
```

```
ggplot(df) +
  aes(x = comprendre_un_texte_en_français, y = Sexe) +
  geom_count(colour="hotpink", alpha=.2) +
  xlab("Peut comprendre un texte en français") +
  ylab("Sexe")+labs(size="effectifs")
```



Le premier tableau montre que la majeure partie de notre population ne sait pas lire le français aussi bien les femmes que les hommes. Ce constat est pratiquement le même quand on regarde ceux qui savent écrire ou comprendre la langue de molière.

### III.2 Accès à l'école suivant le sexe

```
# Croisons sexe et fait_une_ecole_formelle
```

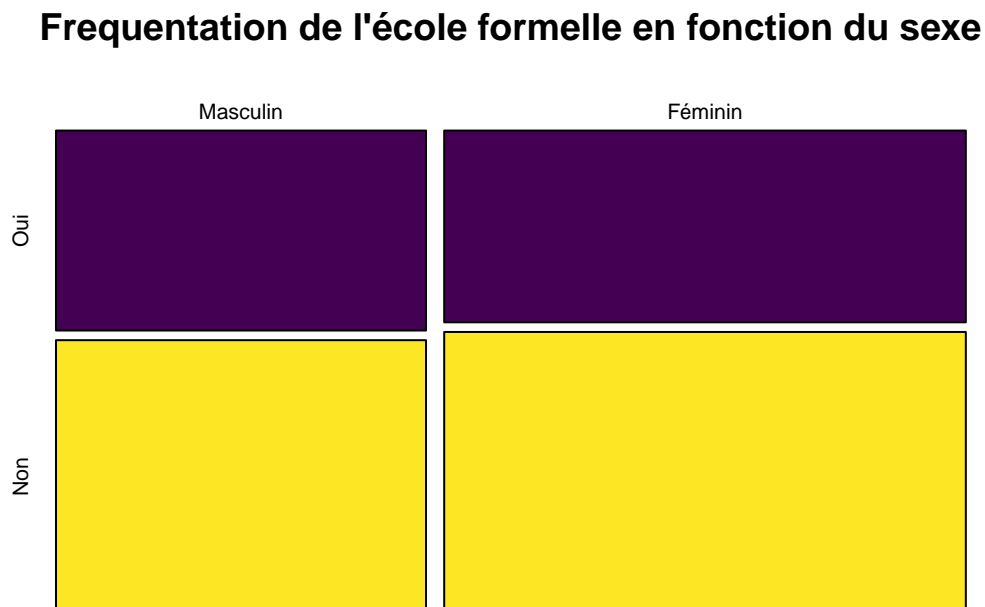
```
tab11 <- table(df$Sexe,df$Fait_une_ecole_formelles)
print(tab11)
```

```
##
##           Oui  Non
## Masculin 2422 3312
## Féminin  3274 4809
```

```
#Representation
library(viridis)
```

```
## Le chargement a nécessité le package : viridisLite
```

```
mosaicplot(tab11,main = "Frequentation de l'école formelle en fonction du sexe", col = viridis(2) )
```





```
chisq.test(tab11) #test d KHI2
```

```
##  
## Pearson's Chi-squared test with Yates' continuity correction  
##  
## data: tab11  
## X-squared = 4.0938, df = 1, p-value = 0.04304
```

Il ressort que les personnes qui n'ont pas accès aux écoles formelles sont en nombre important comparé à ceux qui en ont aussi bien chez les filles que chez les garçons. Ces faibles divergences entre le groupe des filles et celui des garçons est confirmée par un test de Khi2 qui rejette l'existence de dépendance entre ces deux variables.

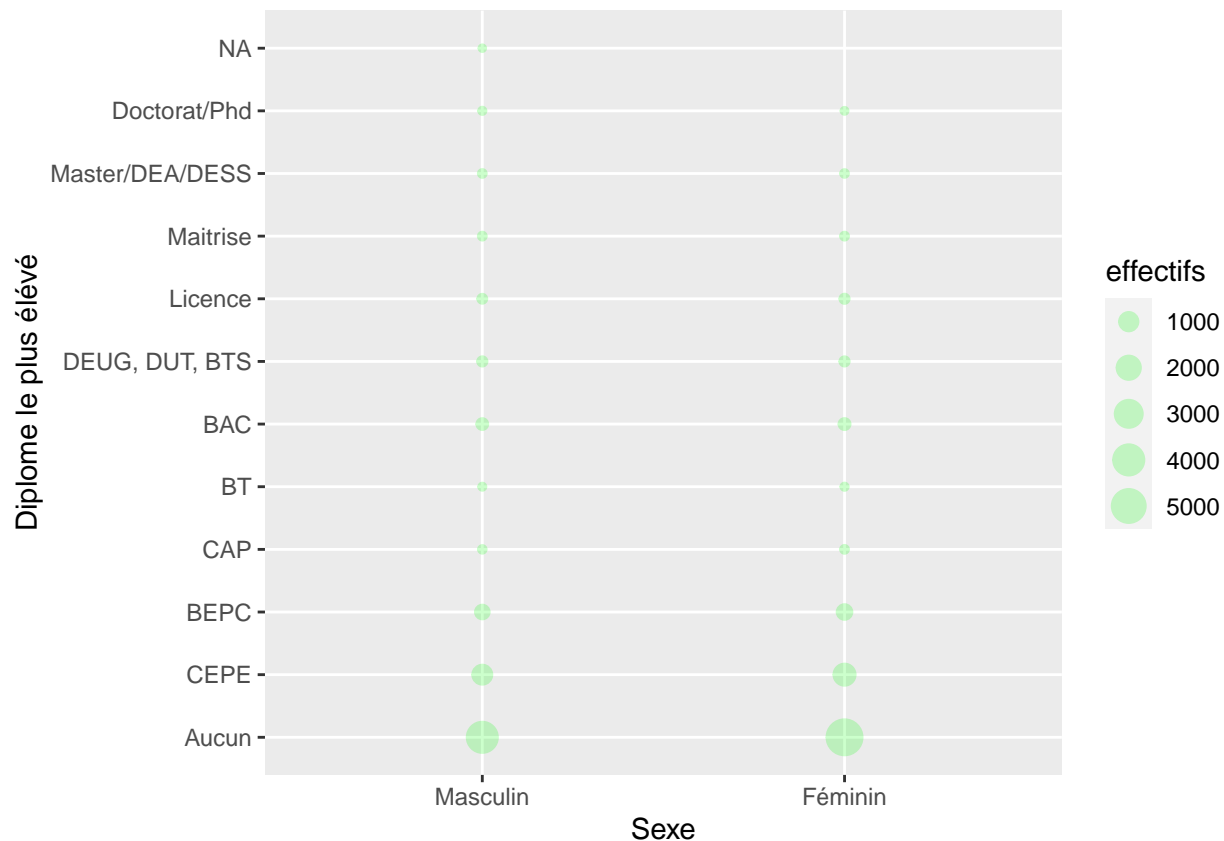
### III.3 Niveau d'éducation suivant le sexe

```
#croisons sexe et diplome le plus ?lev?
```

```
tab12 <- table(df$Diplome_le_plus_eleve,df$Sexe)  
print(tab12)
```

```
##  
##           Masculin Féminin  
##   Aucun           3983    5799  
##   CEPE             1111    1522  
##   BEPC              367     474  
##   CAP                9      13  
##   BT                 3       4  
##   BAC              134     142  
##   DEUG, DUT, BTS    54      52  
##   Licence           41      50  
##   Maitrise          14      14  
##   Master/DEA/DESS   13      11  
##   Doctorat/Phd      4       2
```

```
ggplot(df) +  
  aes(x = Sexe, y = Diplome_le_plus_eleve) +  
  geom_count(colour="green", alpha=.2) +  
  xlab("Sexe") +  
  ylab("Diplome le plus élevé")+labs(size="effectifs")
```



En matière de niveaux d'instruction la tendance est la même qu'on soit du côté des hommes que de celui des femmes. La majeure partie de la population n'a aucun diplôme. Et dans l'ensemble des gens qui ont reçu une éducation ont arrêté en ayant que le CEPE. Bref, il y a peu de personnes ayant des diplômes d'études supérieurs.

## IV. Création d'indicateur : Indicateur de maîtrise d'une langue (IML)

L'un des objectifs de notre étude comme mentionné en début de partie 1 est d'évaluer le niveau d'alphabétisation en quantifiant et présentant le niveau de maîtrise d'une langue. Pour ce faire nous allons élaborer un indice synthétique de maîtrise de langue qui sera une moyenne géométrique d'indices préalablement calculés.

### IV.1 Etape 1 : indice simple

La première étape pour élaborer notre indicateur consiste à calculer des indices simples qui sont la proportion de personnes pouvant lire, comprendre, et écrire un petit texte en une langue.

```
## Etape 1 indice simple = proportion de oui pour les variables langue
langue <- c("Peut_lire_français" , "Peut_lire_la_langue_locale", "Peut_lire_une_autre_langue",
            "Peut_ecrire_français", "Peut_ecrire_la_langue_locale", "Peut_ecrire_une_autre_langue",
            "comprendre_un_texte_en_français", "comprendre_texte_langue_locale", "comprendre_texte_en_au")
indices_s <- c()
```

```
for (x in langue) {
  y <- data.frame(prop.table(table(df[x])))
  indices_s <- c( y[which(y[x] == "Oui" | y[x] == 1 ),2],indices_s)
}
Indice <- data.frame(indices_s)
row.names(Indice) <- langue
```

## IV.2 Etape 2 : IML pour chaque groupe de langue

Une fois les indices simple calculés, la prochaine étape consiste à calculer un indice synthétique pour chaque groupe de langue qui n'est autre que la moyenne géométrique des proportions de personnes pouvant lire, comprendre et écrire cette (ces) langues. Cet indicateur sera nommé Indice de Maîtrise de la langue IML en question. Dans notre cas nous parlerons d'indice de maîtrise du français, de la langue locale, des autres langues.

```
## Etape 2 : IML pour chacune des langues
## IML = moyenne géométrique (indices simples de l'étape 1)
### Regrouper les variables par langue
Francais <- c("Peut_lire_français","Peut_ecrire_français","comprendre_un_texte_en_français")
Langue_locale <- c("Peut_lire_la_langue_locale","Peut_ecrire_la_langue_locale","comprendre_texte_langue")
Autre_langues <- c("Peut_lire_une_autre_langue","Peut_ecrire_une_autre_langue","comprendre_texte_en_autre_langue")

### calcul de l'IML pour chaque groupe de langue
library(forestmangr)
IML <- data.frame (c(pow(Indice[Francais[1],]* Indice[Francais[2],]*Indice[Francais[3],],1/length(Francais)),
                    pow(Indice[Langue_locale[1],]*Indice[Langue_locale[2],]*Indice[Langue_locale[3],],1/length(Langue_locale)),
                    pow(Indice[Autre_langues[1],]*Indice[Autre_langues[2],]*Indice[Autre_langues[3],],1/length(Autre_langues)))

rownames(IML) <- c("Français", "Langue locale", "Autre langues")
colnames(IML) <- "Indice_de_maîtrise_de_langue"
```

## IV. Etape 3 : Indice Globale de Maîtrise de langue

La troisième et dernière étape consiste à quantifier le niveau de maîtrise d'au moins une langue à travers l'Indice Globale de Maîtrise de langue IGML. Cet indice se calcule comme le précédent IML à la différence qu'ici les indices simples sont remplacés par les IML précédemment calculés.

```
## Etape 3 : Indice globale de maîtrise de langue IGML
## IGML = moyenne géométrique (IML de l'étape 2)
IGML <- pow(IML[1,]*IML[2,]*IML[3,], 1/length(IML))
print(IGML)
```

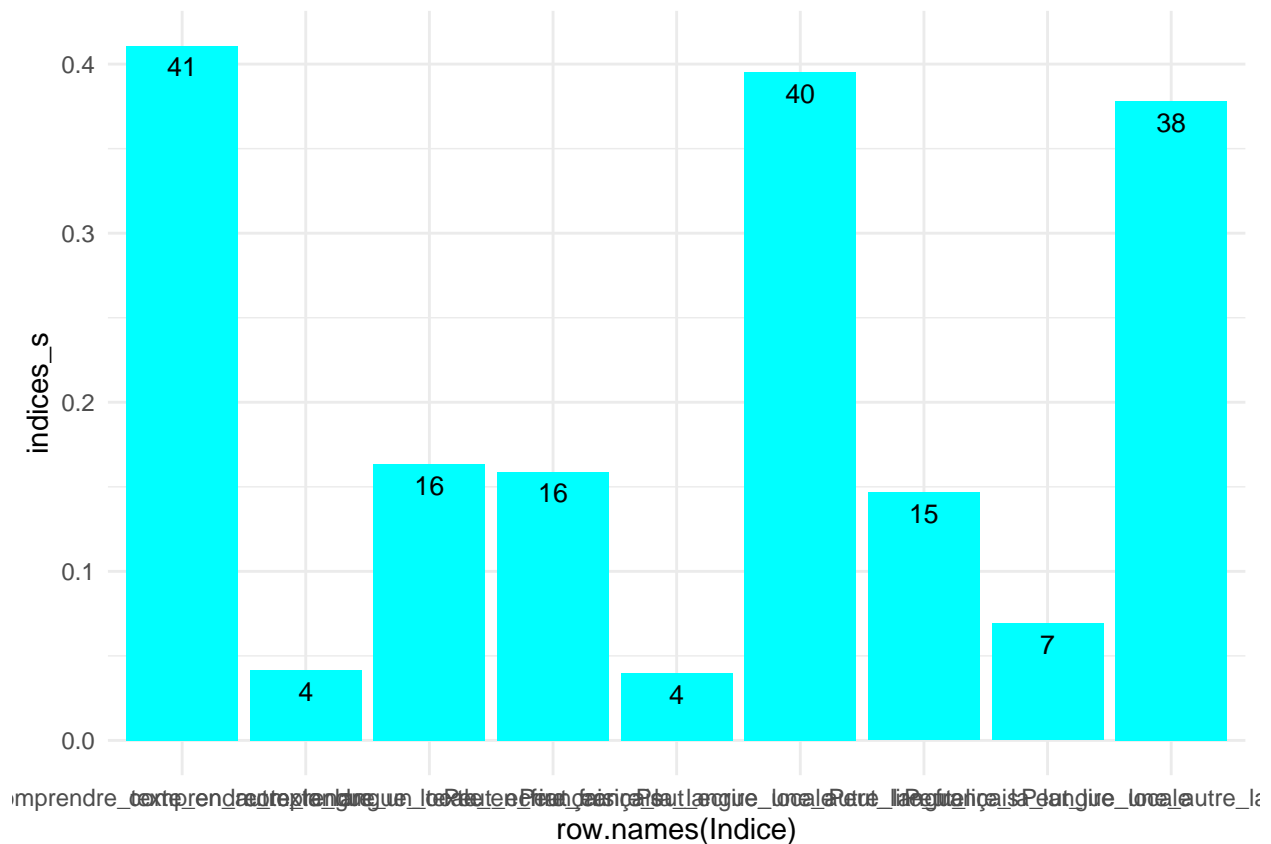
```
## [1] 0.002985561
```

## IV.4 Interprétation de l'indice

### IV.4.1 Indice simple

Les indices simples calculé dans la phase 1 ne sont rien d'autre que la part de la population qui sait lire, écrire ou comprendre un texte dans une langue spécifique

```
### visualisation avec ggplot2
ggplot(data=Indice, aes(x= row.names(Indice),y=indices_s)) +
  geom_bar(stat = "identity",fill="cyan")+
  geom_text(aes(label= round(indices_s*100)), vjust=1.6, color="black", size=3.5)+
  theme_minimal()
```



Le graphique ci-dessus montre que la population étudiée comprend plus de textes en d'autres langues qu'en français.

#### IV.4.2 IML

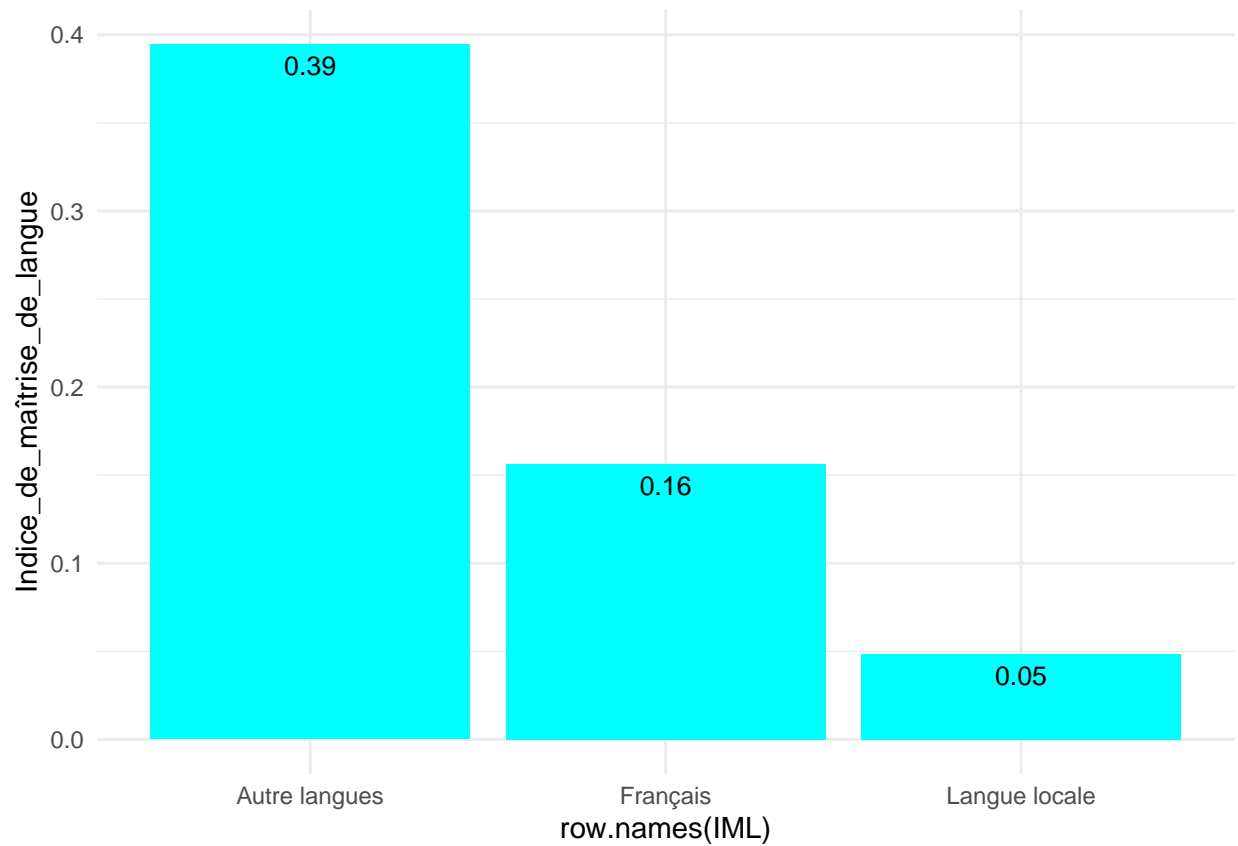
la population qui sait lire, écrire et comprendre un texte dans une langue spécifique

A partir de ces valeurs nous allons définir des intervalles d'interprétation de l'IML:

- IML inférieur à 0.25 : très mauvaise maîtrise de la langue
- IML compris entre 0.25 et 0.5 : mauvaise maîtrise de la langue
- IML compris entre 0.5 et 0.75 : bonne maîtrise de la langue
- IML supérieur à 0.75 : très bonne maîtrise de la langue

```
### visualisation avec ggplot2
ggplot(data=IML, aes(x= row.names(IML),y=Indice_de_maîtrise_de_langue)) +
  geom_bar(stat = "identity",fill="cyan")+
  theme_minimal()
```

```
geom_text(aes(label= round(IML[,1],2)), vjust=1.6, color="black", size=3.5)+
theme_minimal()
```



Le graphique ci-dessus montre que sur notre échantillon, on a en général une mauvaise maîtrise des langues.

#### IV.4.3 IGML

l'IGML peut être perçue comme la part de la population qui sait lire, écrire ou comprendre un texte dans une langue. Et il est très faible pour notre échantillon.

```
print(IGML)
```

```
## [1] 0.002985561
```