

Analyse de l'Anémie comme indicateur de malnutrition chez les enfants ghanéens

2023-11-08

Problématique:

La santé et le bien-être des enfants sont des indicateurs essentiels du développement durable d'une société. Au cœur de la santé infantile se trouve la nutrition, un domaine influencé par une multitude de facteurs (l'environnement familial à l'accèsibilité des ressources, les pratiques éducatives). L'objectif principal de l'enquête Démographique et de Santé (EDHS) de 2014 au Ghana, était de générer des informations fiables sur la planification familiale, la mortalité infanto-juvénile, la santé maternelle et infantile, ainsi que la malnutrition infantile. C'est dans cette perspective que nous nous sommes intéressés à la malnutrition au Ghana, en choisissant l'anémie comme indice. Quels sont les facteurs susceptibles d'influencer l'anémie chez les enfants de moins de 5 ans, tels que le niveau d'éducation de la mère, la richesse du ménage, l'environnement et la prévalence de la maladie du paludisme ?

Pour apporter des éléments de réponse, nous allons parcourir les étapes ci-dessous :

- Prétraitement des données
- Analyse Univariée
- Analyse bivariée
- ACM

Prétraitement des données

Lecture de la Table

```
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

library(tidyverse)
```

```
## — Attaching core tidyverse packages ————— tidyverse 2.0.0 —
## ✓forcats 1.0.0 ✓readr 2.1.4
## ✓ggplot2 3.4.3 ✓stringr 1.5.0
## ✓lubridate 1.9.2 ✓tibble 3.2.1
## ✓purrr 1.0.2 ✓tidyr 1.3.0

## — Conflicts ————— tidyverse_conflicts() —
## X dplyr::filter() masks stats::filter()
## X dplyr::lag() masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(haven)
```

```
menage <- read_sas("GHHR72FL.SAS7BDAT")
```

Quelques variables intéressantes pour notre études:

- Anemia (HA57)
- Mother's highest educational (HC61)
- Food cooked in house (HV241)
- Type of cooking (HV226)
- Region (HV024)
- Number of children 5 and under (HV014)
- Type of place of residence (HV025)
- Source of drinking water (HV201)
- Child's age in months (HC1)

Selection des enfants

Pour pouvoir répondre à notre problématique, il nous a fallu transformer la table d'origine. En effet, les lignes de la table d'origine correspondaient aux ménages, alors que notre étude est basée sur les enfants. Pour ce faire, nous avons créé un nouveau tableau nommé "enfant", dont chaque ligne correspond à un enfant. Dans ce nouveau tableau, chaque enfant a sa propre valeur pour des variables individuelles, telles que le niveau d'anémie, le niveau maximal d'études atteint par la mère, l'âge de l'enfant, et bien d'autres variables communes avec les enfants du même ménage.

Dans un premier temps, nous devons déterminer le nombre d'enfants de moins de 5 ans dans la table "ménage". Pour cela, nous avons décidé d'observer la variable HC1 (âge des enfants), composée de 7 colonnes (HC1_1, ... HC1_7). En utilisant le code ci-dessus, nous avons constaté que l'âge minimum est de 0 mois et l'âge maximum est de 59 mois (5 ans), et que la colonne HC1_7 est vide.

résumé des colonnes HC1

```
summary(select(menage, HC1_1:HC1_7))
```

```

##      HC1_1          HC1_2          HC1_3          HC1_4
## Min.   : 0.00   Min.   : 0.00   Min.   : 0.00   Min.   : 1.00
## 1st Qu.:21.00  1st Qu.: 7.00  1st Qu.: 5.00  1st Qu.: 8.25
## Median :35.00  Median :16.00  Median :12.00  Median :14.50
## Mean    :33.49  Mean    :18.11  Mean    :16.89  Mean    :18.68
## 3rd Qu.:47.00  3rd Qu.:26.00  3rd Qu.:23.00  3rd Qu.:28.50
## Max.    :59.00  Max.    :59.00  Max.    :58.00  Max.    :48.00
## NA's    :9521   NA's    :11086  NA's    :11710  NA's    :11813
##      HC1_5          HC1_6          HC1_7
## Min.   :43.00   Min.   : 1.0    Min.   : NA
## 1st Qu.:46.00  1st Qu.: 3.0    1st Qu.: NA
## Median :49.00  Median : 5.0    Median : NA
## Mean   :47.33  Mean   :10.0    Mean   :NaN
## 3rd Qu.:49.50  3rd Qu.:14.5    3rd Qu.: NA
## Max.   :50.00  Max.   :24.0    Max.   : NA
## NA's   :11832   NA's   :11832   NA's   :11835

```

Pour compter le nombre d'enfant, on a compté le nombre de lignes qui contiennent des valeurs pour les 6 colonnes HC1_1, jusqu'à HC1_6. Au total on a trouvé 3238.

Nombre des enfants

```

data1 <- menage %>%
  filter(!is.na(HC1_1))
data2<- menage %>%
  filter(!is.na(HC1_2))
data3<- menage %>%
  filter(!is.na(HC1_3))
data4<- menage %>%
  filter(!is.na(HC1_4))
data5<- menage %>%
  filter(!is.na(HC1_5))
data6<- menage %>%
  filter(!is.na(HC1_6))
data7<- menage %>%
  filter(!is.na(HC1_7))
nbrenf <- nrow(rbind(data1,data2,data3,data4,data5,data6,data7))
nbrenf

```

```
## [1] 3238
```

Après observation, nous avons remarqué que la variable que nous avons décidé d'examiner, HA57 (Anémie des enfants), possède 6 colonnes dans notre tableau, à savoir HA57_1, HA57_2, jusqu'à HA57_7. En d'autres termes, nous avons au maximum 7 enfants par ménage, et les colonnes mentionnées précédemment correspondent respectivement au premier, deuxième jusqu'au septième enfant. Notre objectif est d'obtenir le niveau d'anémie pour chaque enfant de moins de 5 ans.

Pour sélectionner les premiers enfants de chaque ménage, nous avons conservé les lignes de la table ménage qui n'ont pas de valeurs manquantes (NA) pour la variable HA57_1. Il est important de noter que les lignes contenant des valeurs manquantes (222 lignes) correspondent aux ménages qui n'ont pas d'enfant de moins de 5 ans ou dont le niveau d'anémie de cet enfant est inconnu.

Ensuite, pour la variable HC61 (Niveau d'étude de la mère), qui est composée de 7 colonnes également, nous avons appliqué la même procédure que celle utilisée pour HA57_1, mais cette fois-ci avec HC61_1.

Pour trouver les autres enfants, nous avons effectué de manière successive les opérations précédentes pour les variables HA57 (Anémie) et HC61 (niveau d'étude de la mère) qui correspondent aux enfants supplémentaires.

Au final, nous avons identifié 2175 enfants pour lesquels nous disposons à la fois du niveau d'anémie et du niveau d'éducation de la mère. Les autres variables, telles que l'indice de richesse, la région et l'électricité, sont des variables partagées par tous les membres du ménage et ne présentent pas de valeurs manquantes (NA) pour les individus de notre population. Par conséquent, aucune étape de prétraitement n'est nécessaire pour ces variables.

```
summary(data1$HA57_1)
```

```
##      Min. 1st Qu. Median     Mean 3rd Qu.    Max.    NA's
## 1.000   3.000  4.000   3.483  4.000   4.000    222
```

```
summary(data2$HC61_1)
```

```
##      Min. 1st Qu. Median     Mean 3rd Qu.    Max.    NA's
## 0.000   0.000  1.000   1.014  2.000   3.000    50
```

Selection des enfants

```
# Les premiers enfants

enf1 <- filter(menage, !is.na(HA57_1))

enf1<- filter(enf1, !is.na(HC61_1))

enf1$HA57 = enf1$HA57_1
enf1$HC61=enf1$HC61_1

enf1 <-filter(enf1,HC1_1<=60)

enf1$HC1=enf1$HC1_1

enf1$HC27=enf1$HC27_1

#####
##

# Les deuxièmes enfants
enf2 <- enf1 %>%
  filter(!is.na(HA57_2), !is.na(HC61_2))

enf2$HA57= enf2$HA57_2
enf2$HC61 <-enf2$HC61_2

enf2 <-filter(enf2,HC1_2<=60)
enf2$HC1=enf2$HC1_2

enf2$HC27=enf2$HC27_2

#####
# Les troisièmes enfants

enf3 <- enf2 %>%
  filter(!is.na(HA57_3), !is.na(HC61_3))

enf3$HA57 = enf3$HA57_3
enf3$HC61 = enf3$HC61_3

enf3 <-filter(enf3,HC1_3<=60)
enf3$HC1=enf3$HC1_3

enf3$HC27=enf3$HC27_3

#####
# Les quatrièmes enfants

enf4 <- enf3 %>%
```

```
filter(!is.na(HA57_4), !is.na(HC61_4))

enf4$HA57 = enf4$HA57_4
enf4$HC61 = enf4$HC61_4

enf4 <- filter(enf4, HC1_4<=60)
enf4$HC1=enf4$HC1_4

enf4$HC27=enf4$HC27_4

# combinaisons des 4 tableaux
enfant <- rbind(enf1,enf2,enf3,enf4)

enf<-select (enfant, HA57, HC61, HV270, HV024, HC1, HC27, HV025, HHID, HV005, HV009, HV012, HV013, HV014, HV024, HV025, HV201, HV270, HV204:HV228, HV243A, SH110L, SH110P, SH110Q, SH110R, SH110S)

#colnames(enf) <- c("HA57", "HC61", "HV270", "HV024", "HC1", "HC27", "HV025", "HHID", "HV005", "HV009", "HV012", "#HV013", "HV014", "HV024", "HV025", "HV201", "HV270", "HV204", "HV205", "HV206", "HV207", "HV208", "HV209", "#HV210", "HV211", "HV212", "HV213", "HV214", "HV215", "HV216", "HV217", "HV218", "HV219", "HV220", "HV221", "#HV222", "HV223", "HV224", "HV225", "HV226", "HV227", "HV228", "HV243A", "SH110L", "SH110P", "SH110Q", "#SH110R", "SH110S")
```

Conversion des variables

```

enf$cHA57 <- as.factor(enf$HA57)
enf$cHC27 <- as.factor(enf$HC27)
enf$cHC61 <- as.factor(enf$HC61)
enf$cHV024 <- as.factor(enf$HV024)
enf$cHV270 <- as.factor(enf$HV270)
enf$cHV025 <- as.factor(enf$HV025)
enf$cHV024 <- as.factor(enf$HV024)

enf$cHV201 <- as.factor(enf$HV201)
enf$cHV204 <- as.factor(enf$HV204)
enf$cHV205 <- as.factor(enf$HV205)
enf$cHV206 <- as.factor(enf$HV206)
enf$cHV207 <- as.factor(enf$HV207)
enf$cHV208 <- as.factor(enf$HV208)
enf$cHV209 <- as.factor(enf$HV209)
enf$cHV210 <- as.factor(enf$HV210)
enf$cHV211 <- as.factor(enf$HV211)
enf$cHV212 <- as.factor(enf$HV212)
enf$cHV213 <- as.factor(enf$HV213)
enf$cHV214 <- as.factor(enf$HV214)
enf$cHV215 <- as.factor(enf$HV215)
enf$cHV216 <- as.factor(enf$HV216)
enf$cHV217 <- as.factor(enf$HV217)
enf$cHV218 <- as.factor(enf$HV218)
enf$cHV219 <- as.factor(enf$HV219)
enf$cHV220 <- as.factor(enf$HV220)
enf$cHV221 <- as.factor(enf$HV221)
enf$cHV225 <- as.factor(enf$HV225)
enf$cHV226 <- as.factor(enf$HV226)
enf$cHV227 <- as.factor(enf$HV227)
enf$cHV228 <- as.factor(enf$HV228)

enf$cHV243A <- as.factor(enf$HV243A)
enf$cHV270 <- as.factor(enf$HV270)
enf$cSH110L <- as.factor(enf$SH110L)
enf$cSH110P <- as.factor(enf$SH110P)
enf$cSH110Q <- as.factor(enf$SH110Q)
enf$cSH110R <- as.factor(enf$SH110R)
enf$cSH110S <- as.factor(enf$SH110S)

```

Analyse Statistique:

Etude univariée

Il s'agit d'une étape cruciale en analyse statistique. Cette étape va servir d'une base solide pour faire des analyses statistiques plus avancées et à tirer des conclusions plus informées dans le cadre de notre travail.

Nous allons réaliser le tris à plat des variables pertinentes pour notre étude.

- Âge des enfants en mois (HC1)
- Niveau d'étude de la mère (HC61)
- Présence de moustiquaire autour des lits (HV227)

- Region (HV024)
- milieu du résidence (rural ou urbain) (HV025)
- Source de l'eau (HV201)
- Electricité
- Source de richesse
- Anémie (HA57)

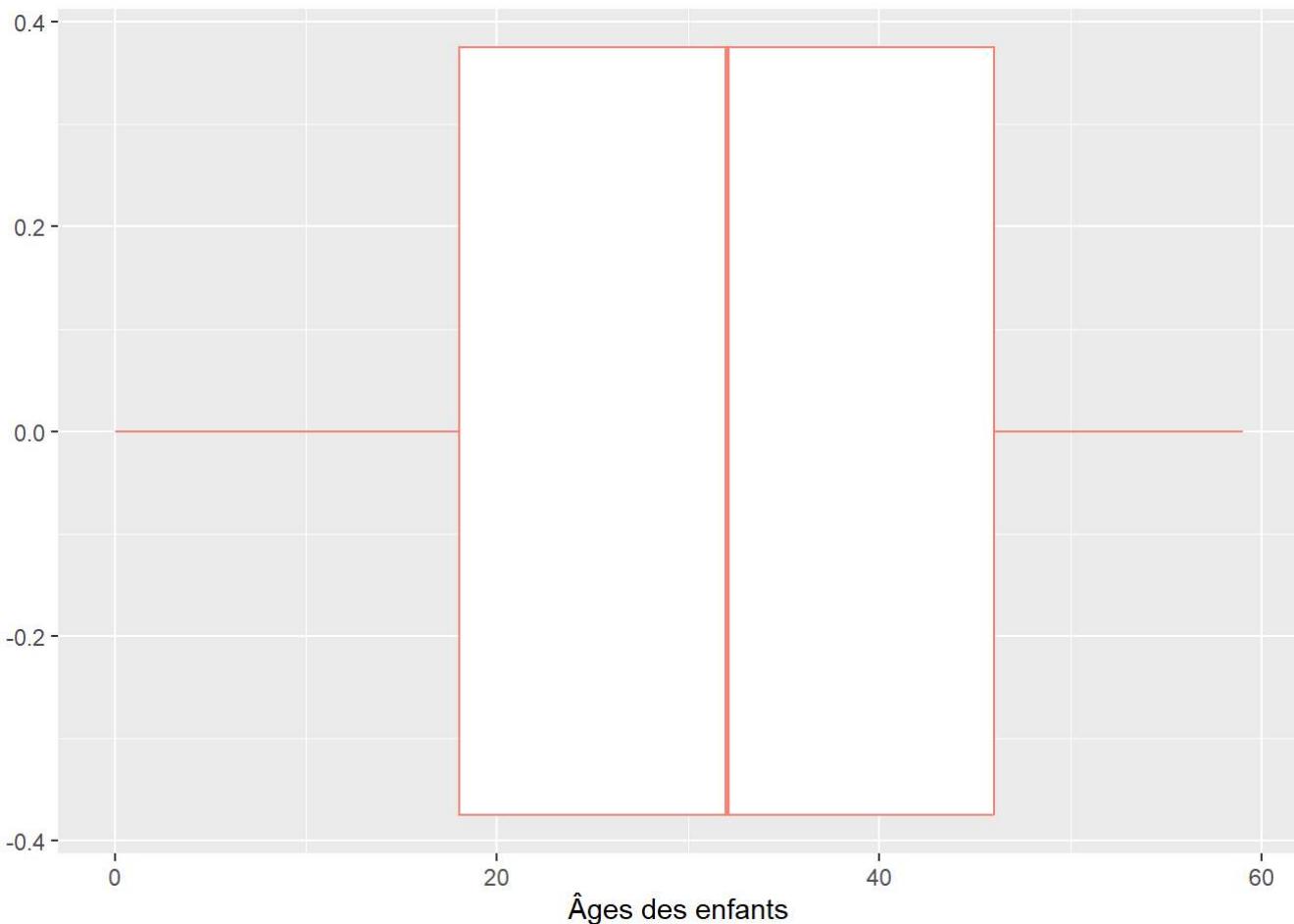
Etude de la variable âge des enfants moins de 5 ans en mois (HC1)

Dans un premier temps, nous avons décidé d'observer la distribution de l'âge de notre population, à savoir les enfants de moins de 5 ans. Cette variable nous permettra de déterminer quelle tranche d'âge est la plus touchée par l'anémie.

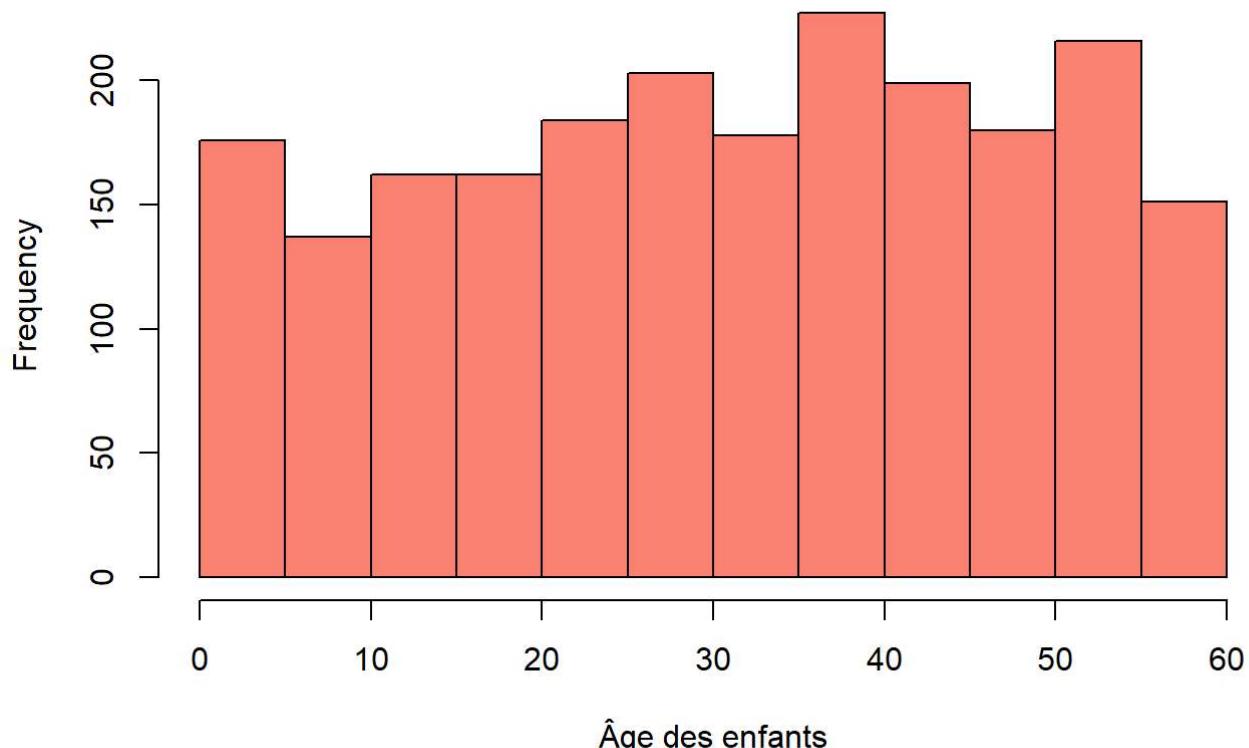
Pour ce faire, nous avons tracé le boxplot et l'histogramme de cette variable, présentés dans les figures ci-dessous.

```
library(ggplot2)

ggplot(enf, aes(x=HC1))+geom_boxplot(color="salmon")+labs(x="Âges des enfants")
```



```
hist(enf$HC1,xlab= "Âge des enfants",main="", col='salmon')
```



D'après l'histogramme, l'âge des enfants de moins de 5 ans est bien réparti. Pour chaque tranche d'âge, nous avons presque le même nombre d'enfants. Cette variable présente une distribution uniforme. Une telle répartition aura un impact sur les interprétations de notre étude, car les résultats sur l'anémie des enfants ne seront pas restrictifs à une tranche d'âge spécifique.

En ce qui concerne le boxplot, nous remarquons une symétrie dans la distribution de l'âge des enfants. Cela signifie que le nombre de valeurs inférieures et supérieures à la médiane (32 mois) est à peu près égal. Nous notons également l'absence de valeurs aberrantes en dehors des moustaches, ce qui confirme la bonne distribution des données.

Etude de la variable Avoir une moustiquaire (HV227)

L'utilisation de moustiquaires est souvent associée à la prévention des maladies transmises par les moustiques, notamment le paludisme. Dans le contexte de l'étude sur l'anémie chez les enfants, la variable "moustiquaire" peut revêtir une importance particulière, car les moustiques infectieux peuvent être responsables de la transmission de maladies pouvant causer ou aggraver l'anémie chez l'enfant.

```
library(labelled)
```

```
## Warning: le package 'labelled' a été compilé avec la version R 4.3.2
```

```
library(openxlsx)
```

```
## Warning: le package 'openxlsx' a été compilé avec la version R 4.3.2
```

```

# création de label pour la variable Anémie
attr(enf$HA57, "label") <- "Anémie"

# effectifs
var_mousti <- enf %>% group_by(cHV227)%>%
  summarize( effectifs=n())
var_mousti$eff_cum <-cumsum(var_mousti$effectifs)

# Fréquences
var_mousti$pourcent <-round(prop.table(var_mousti$effectifs)*100,2)
var_mousti$pour_cum <- cumsum(var_mousti$pourcent)

# Modalités
mousti <- c("Non","Oui")
var_mousti$moustiquaire <- mousti

# Noms des colonnes
colnames(var_mousti)<-c("Moustiquaire", "effectifs","effectifs cumulés","fréquences","fréquences cumulées","presence")

#Exportation de la table
write.xlsx(var_mousti,file="var_mousti.xlsx")

#Affichage du tableau des statistiques de la variable
var_mousti

```

```

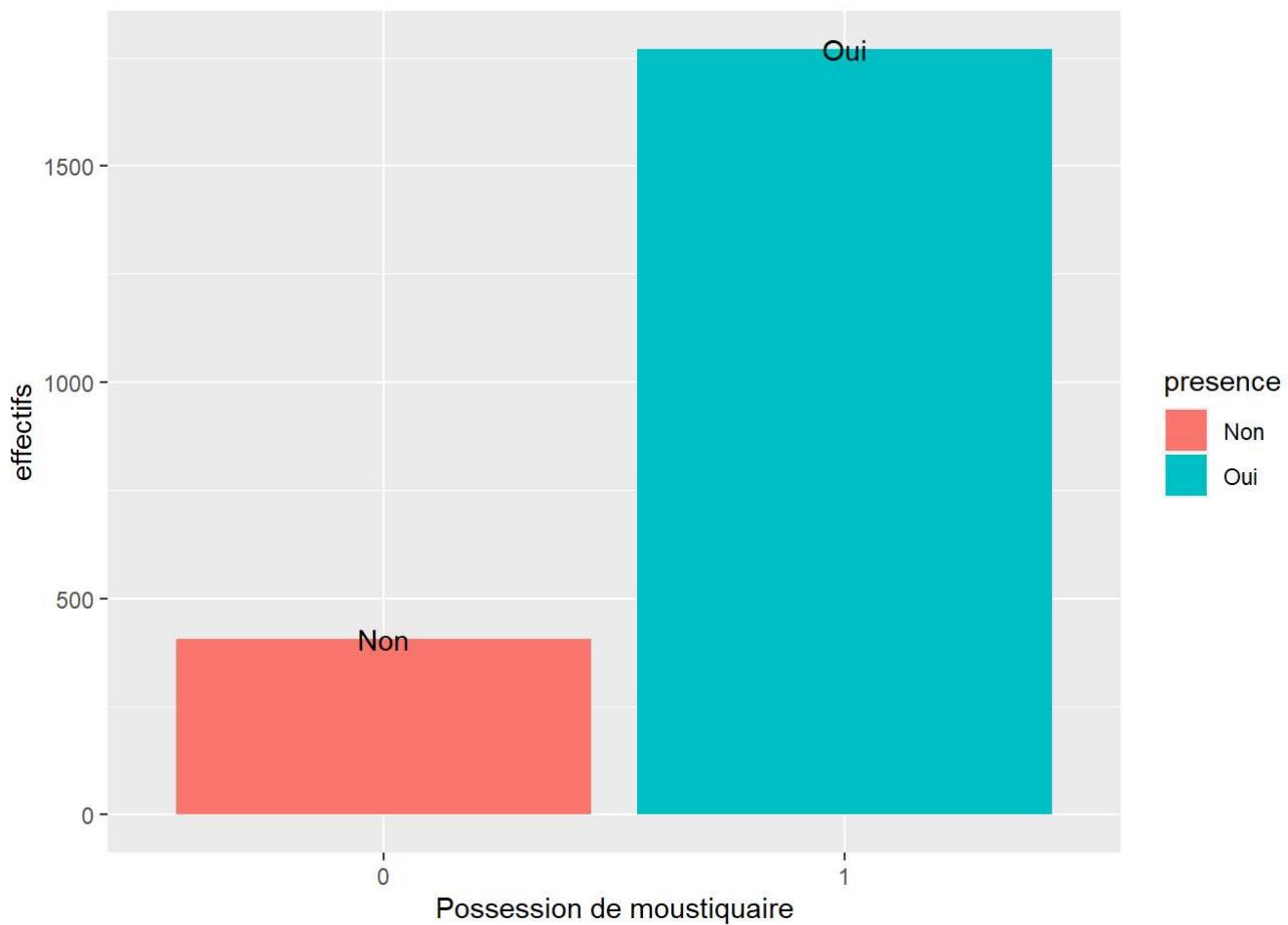
## # A tibble: 2 × 6
##   Moustiquaire effectifs `effectifs cumulés` fréquences `fréquences cumulées`
##   <fct>          <int>                 <int>      <dbl>                <dbl>
## 1 0              406                  406       18.7                18.7
## 2 1             1769                 2175      81.3                100
## # i 1 more variable: presence <chr>

```

```

# Histogramme
ggplot(var_mousti,aes(x=Moustiquaire,y=effectifs ,fill=presence))+ geom_bar(stat = "identity")+
  labs(x="Possession de moustiquaire", y="effectifs")+ geom_text(label= mousti)

```



L'histogramme ci-dessus montre qu'environ 80% des ménages où il y a des enfants de moins de 5 ans possèdent des moustiquaires autour des lits. Dans notre étude, nous avons l'intuition que les enfants qui sont dans des ménages sans moustiquaires autour des lits sont susceptibles d'être plus affectés par l'anémie.

Etude de la variable Niveau d'éducation de la mère (HC61)

Cette variable semble être très importante dans notre étude. En effet, nous pensons qu'une mère avec un niveau d'éducation élevé est plus informée sur les risques des maladies de son enfant et donc veillera à sa santé et à bien le nourrir.

```

# Effectifs
var_etud <- enf %>% group_by(chC61)%>%
  summarize( effectifs=n())
var_etud$eff_cum <-cumsum(var_etud$effectifs)

# Fréquences
var_etud$pourcent <-round(prop.table(var_etud$effectifs)*100,2)
var_etud$pour_cum <- cumsum(var_etud$pourcent)

# Modalités
etude <- c("Pas d'education", "Primaire", "Sécondaire", "Lycée")
var_etud$etude <- etude

# Noms des colonnes
colnames(var_etud)<-c("Niv_Ed_M", "effectifs","effectifs cumulés","fréquences","fréquences cumulées","niveaux")

#Affichage du tableau des statistiques de la variable
var_etud

```

```

## # A tibble: 4 × 6
##   Niv_Ed_M effectifs `effectifs cumulés` fréquences `fréquences cumulées`
##   <fct>     <int>           <int>      <dbl>            <dbl>
## 1 0          761             761       35.0            35.0
## 2 1          435            1196      20               55.0
## 3 2          902            2098      41.5            96.5
## 4 3          77              2175      3.54            100
## # i 1 more variable: niveaux <chr>

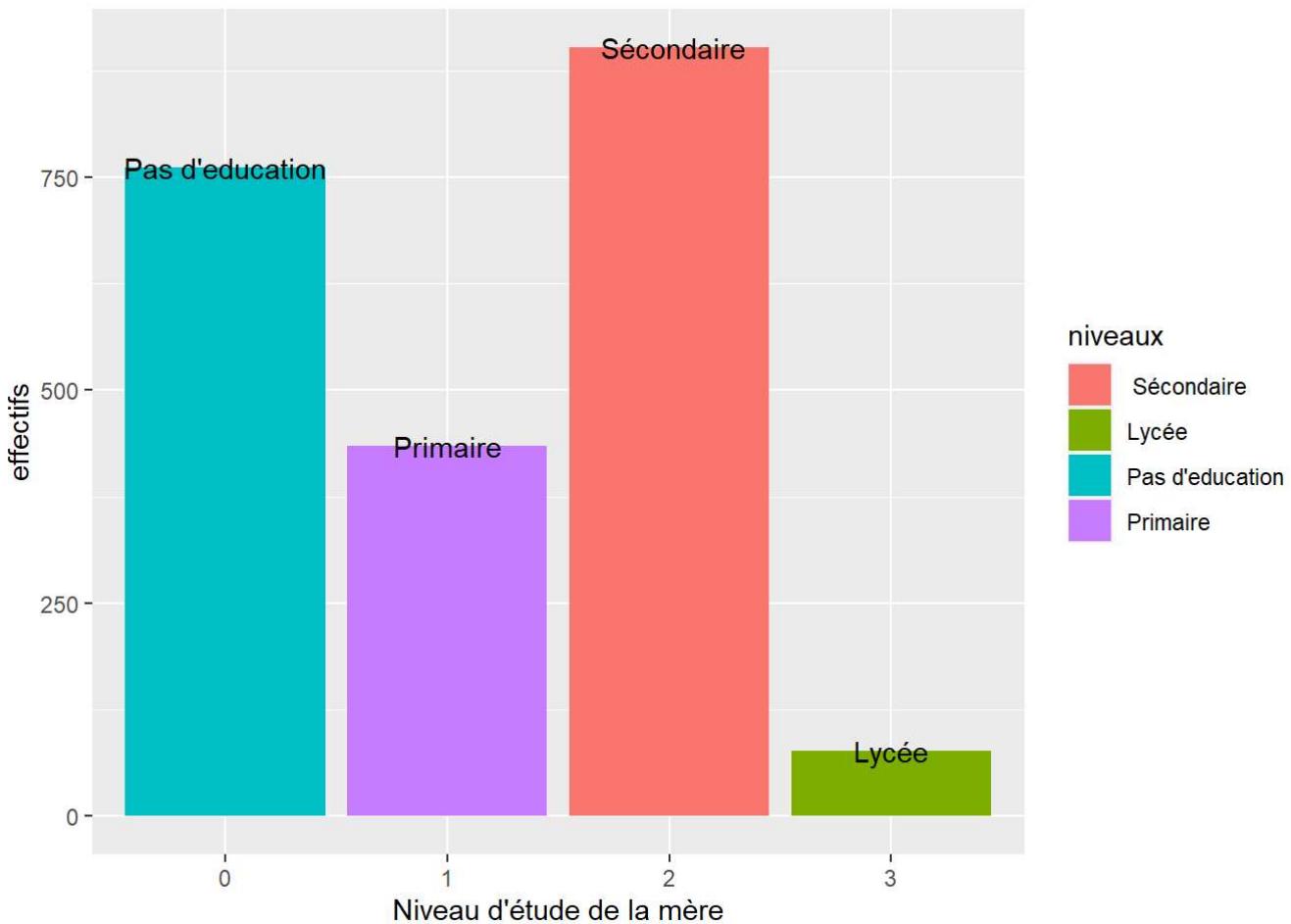
```

```

# Exportation de la table
write.xlsx(var_etud,file="var_etud.xlsx")

# Diagramme
ggplot(var_etud,aes(x=Niv_Ed_M,y= effectifs ,fill=niveaux))+ geom_bar(stat = "identity")+
  labs(x="Niveau d'étude de la mère", y="effectifs") + geom_text(label= etude)

```



```
var_etud
```

```
## # A tibble: 4 × 6
##   Niv_Ed_M effectifs `effectifs cumulés` fréquences `fréquences cumulées`
##   <fct>     <int>           <int>      <dbl>            <dbl>
## 1 0          761             761       35.0            35.0
## 2 1          435             1196      20                55.0
## 3 2          902             2098      41.5            96.5
## 4 3          77              2175      3.54            100
## # i 1 more variable: niveaux <chr>
```

```
#prop.test(table(menage$cHC226), conf.Level = 0.95)
```

En examinant l'histogramme, il est notable que 35% des enfants ont des mères non scolarisées. Par conséquent, la majorité des enfants (65%) ont des mères ayant fait des études.

Plus précisément, les données indiquent que : - 20 % des enfants ont des mères ayant atteint au maximum le niveau primaire, - 41 % des enfants ont des mères ayant atteint le niveau secondaire, - et seulement 3,5 % des enfants ont des mères possédant un niveau d'études lycée.

Etude de la variable Sexe de l'enfant (HC27)

Nous avons décidé d'explorer l'impact de l'anémie sur les enfants, en particulier en examinant si cette condition affecte différemment les deux sexes, à savoir les garçons et les filles. L'étude de cette variable nous aidera à déterminer si l'anémie peut être considérée comme étant plus restrictivement liée à l'un des sexes par rapport à l'autre au sein de la population étudiée.

```

# regroupement et calcul de l'effectif par modalité
var_sex <- enf %>% group_by(cHC27)%>%
  summarize( effectifs=n())
var_sex$eff_cum <-cumsum(var_sex$effectifs)

# Fréquences
var_sex$pourcent <- round(prop.table(var_sex$effectifs)*100,2)
var_sex$pour_cum <- cumsum(var_sex$pourcent)

# Modalités
sexe <- c("garçon", "fille")
var_sex$sexe=sexe

# Noms des colonnes
colnames(var_sex)<-c(" Sexe de l'enfant", "effectifs","effectifs cumulés","fréquences","fréquences cumulées","Sexe")

#Exportation de la table
write.xlsx(var_sex,file="var_sex.xlsx")

#Affichage du tableau des statistiques de la variable
var_sex

```

```

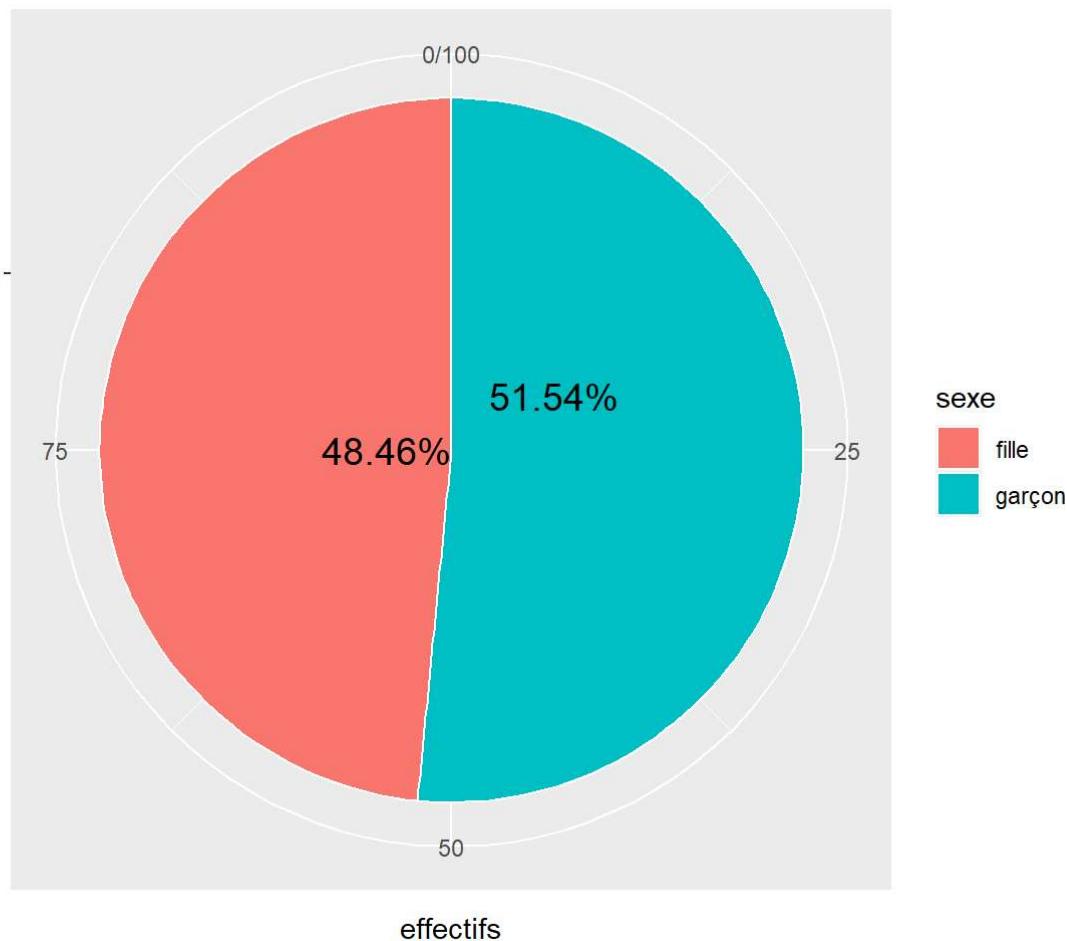
## # A tibble: 2 × 6
##   `Sexe de l'enfant` `effectifs` `effectifs cumulés` `fréquences`
##   <fct>           <int>          <int>        <dbl>
## 1 1                1121            1121        51.5 
## 2 2                1054            2175        48.5 
## # i 2 more variables: `fréquences cumulées` <dbl>, Sexe <chr>

```

```

# Diagramme circulaire
ggplot(var_sex, aes(x = " Sexe de l'enfant", y = fréquences, fill = sexe)) +
  geom_bar(width = 1, stat = "identity", color = "white") +
  coord_polar("y", start = 0) +
  labs(y="effectifs", colour="sexe") +
  geom_text(aes(x = c(1,0.5), y=c(20,0.02),label = paste0(fréquences,"%")),size =5,hjust=1)

```



On remarque à partir du diagramme circulaire que les pourcentages des garçons (52%) et de filles (48%) sont presques les mêmes. On a donc un équilibre de sexe dans notre échantillon ce ce qui peut contribuer à minimiser les biais liés au sexe dans l'analyse . Cela permettra de déterminer s'il existe des différences significatives dans la prévalence de l'anémie entre les deux sexes.

Nous allons réaliser un test d'hypothèses pour savoir si la proportion des garçons dans notre population est de 50%. -Hypothèse nulle (H_0): La probabilité réelle (p) d'obtenir un garçon est égale à 0,5. -Hypothèse alternative (H_1): La probabilité réelle (p) d'obtenir un garçon n'est pas égale à 0,5.

```
# prop.test
(taille_ech <- nrow(enf))

## [1] 2175

(nbr_garçon <- nrow(filter(enf, cHC27==1)))

## [1] 1121

res_test <- prop.test(nbr_garçon,taille_ech,0.5)
res_test
```

```

## 
## 1-sample proportions test with continuity correction
##
## data: nbr_garçon out of taille_ech, null probability 0.5
## X-squared = 2.0028, df = 1, p-value = 0.157
## alternative hypothesis: true p is not equal to 0.5
## 95 percent confidence interval:
## 0.4941608 0.5365889
## sample estimates:
##          p
## 0.5154023

```

Le test ci-dessus, nous donne une p-valeur de 15%. Puisque le p-value est supérieur au seuil de significativité courant de 0,05, on ne rejette pas l'hypothèse nulle. Cependant, on ne peut pas conclure que la probabilité d'obtenir un garçon est exactement de 0,5. L'intervalle de confiance à 95% pour la vraie probabilité (p) d'obtenir un garçon est de 0,494 à 0,537. Cela signifie qu'avec une confiance de 95%, on estime que la proportion réelle de garçons se situe entre ces deux valeurs. La probabilité 0,5 est incluse dans cet intervalle, ce qui est cohérent avec le p-value plus élevé, suggérant que la probabilité réelle pourrait être égale à 0,5. Cependant, il n'y a pas suffisamment de preuves pour rejeter l'hypothèse nulle. Les données ne fournissent pas une base solide pour affirmer que la probabilité d'obtenir un garçon est significativement différente de 0,5.

Etude de la variable Milieu du menage

On a jugé nécessaire de regarder cette variable car l'impact de l'anémie chez les enfants de moins de 5 ans peut avoir un lien avec le milieux. En effet, la malnutrition peut être liées aux disparités socio-économiques entre les milieux urbains et ruraux, les différences d'accès aux soins de santé, à la diversité alimentaire et l'exposition aux risques environnementaux.

```

# regroupement et calcul de l'effectif par modalité
var_zone <- enf %>% group_by(chV025)%>%
  summarize( effectifs=n())
var_zone$eff_cum <-cumsum(var_zone$effectifs)

# Fréquences
var_zone$pourcent<-round(prop.table(var_zone$effectifs)*100,2)
var_zone$pour_cum <- cumsum(var_zone$pourcent)

# Noms des modalités
zone= c("urbaine", "rurale")
var_zone$zone=zone

# Noms des colonnes
colnames(var_zone)<-c("Milieu du ménage", "effectifs","effectifs cumulés","fréquences","fréquences cumulées","zone")

# Exportation de la table
write.xlsx(var_zone,file="var_zone.xlsx")

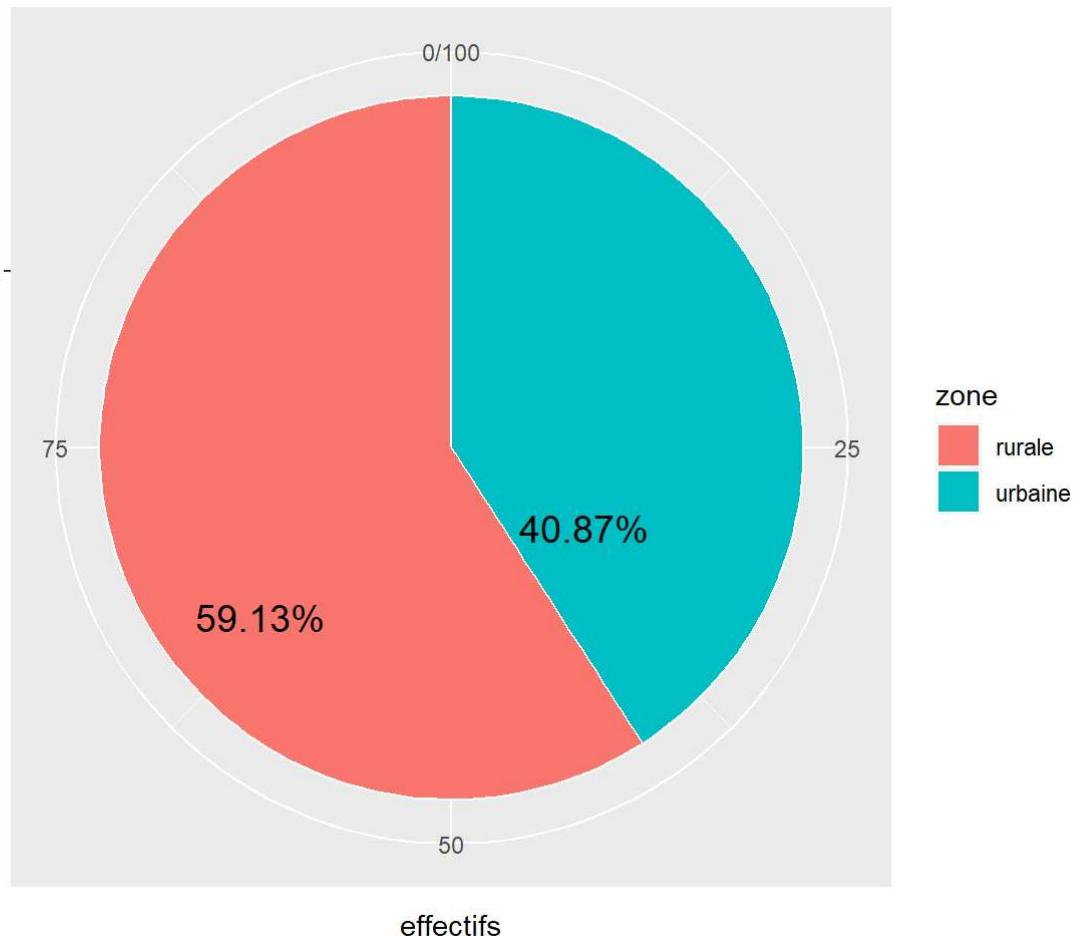
#Affichage du tableau des statistiques de la variable
var_zone

```

```
## # A tibble: 2 × 6
##   `Milieu du ménage` effectifs `effectifs cumulés` fréquences
##   <fct>           <int>           <int>       <dbl>
## 1 1                  889            889        40.9
## 2 2                 1286           2175       59.1
## # i 2 more variables: `fréquences cumulées` <dbl>, zone <chr>
```

```
# diagramme circulaire
```

```
ggplot(var_zone, aes(x = "Milieu du ménage", y = fréquences, fill = zone)) +
  geom_bar(width = 1, stat = "identity", color = "white") +
  coord_polar("y", start = 0) +
  labs(y="effectifs", colour="zone") +
  geom_text(aes(x= c(1.2,1.5),label = paste0(fréquences,"%")), vjust=-4,size = 5)
```



D'après le tableau et le diagramme ci-dessus, on remarque que 1286 enfants moins de 5 ans habitent dans des zones rurales, ce qui représente 59% de l'ensemble des enfants considérés dans l'étude. Et les 41% des enfants sont en zone urbaine.

Etude de la variable Source d'eau (HV201)

Les différentes sources sont: “Canalisée dans le logement”, “Canalisée dans la cour/parcelle”, “Robinet public/fontaine”, “Puits tubulaire ou forage”, “Puits protégé”, “Puits non protégé”, “Source protégée”, “Source non protégée”, “Rivière/barrage/lac/étang/cours d'eau/canal/chenal d'irrigation Eau de pluie”, “Camion-citerne”, “Eau en bouteille”, “Eau en sachet”)

Dans un contexte de nutrition et de santé, l'eau potable est une variable logique à étudier. Les contaminants présents dans l'eau, tels que les agents pathogènes ou les substances chimiques, peuvent avoir des implications directes sur la santé et contribuer à l'anémie.

#Source d'eau HV201 avant regroupement

```

# regroupement et calcul de l'effectif par modalité
var_source1 <- enf %>% group_by(cHV201)%>%
  summarize( effectifs=n())

var_source1$eff_cum <-cumsum(var_source1$effectifs)

# Fréquences
var_source1$pourcent <-round(prop.table(var_source1$effectifs)*100,2)
var_source1$pour_cum <- cumsum(var_source1$pourcent)

# Modalités
source <- c(
  "Canalisée dans le logement",
  "Canalisée dans la cour/parcelle",
  "Robinet public/fontaine",
  "Puits tubulaire ou forage",
  "Puits protégé",
  "Puits non protégé",
  "Source protégée",
  "Source non protégée",
  "Rivière/barrage/lac/étang/cours d'eau",
  "Eau de pluie",
  "Camion-citerne",
  "Eau en bouteille",
  "Eau en sachet")

var_source1$source <- source

# Noms des colonnes
colnames(var_source1)<-c("Source", "effectifs","effectifs cumulés","fréquences","fréquences cumulées","source")

# Exportation de la table
write.xlsx(var_source1,file="var_source1.xlsx")

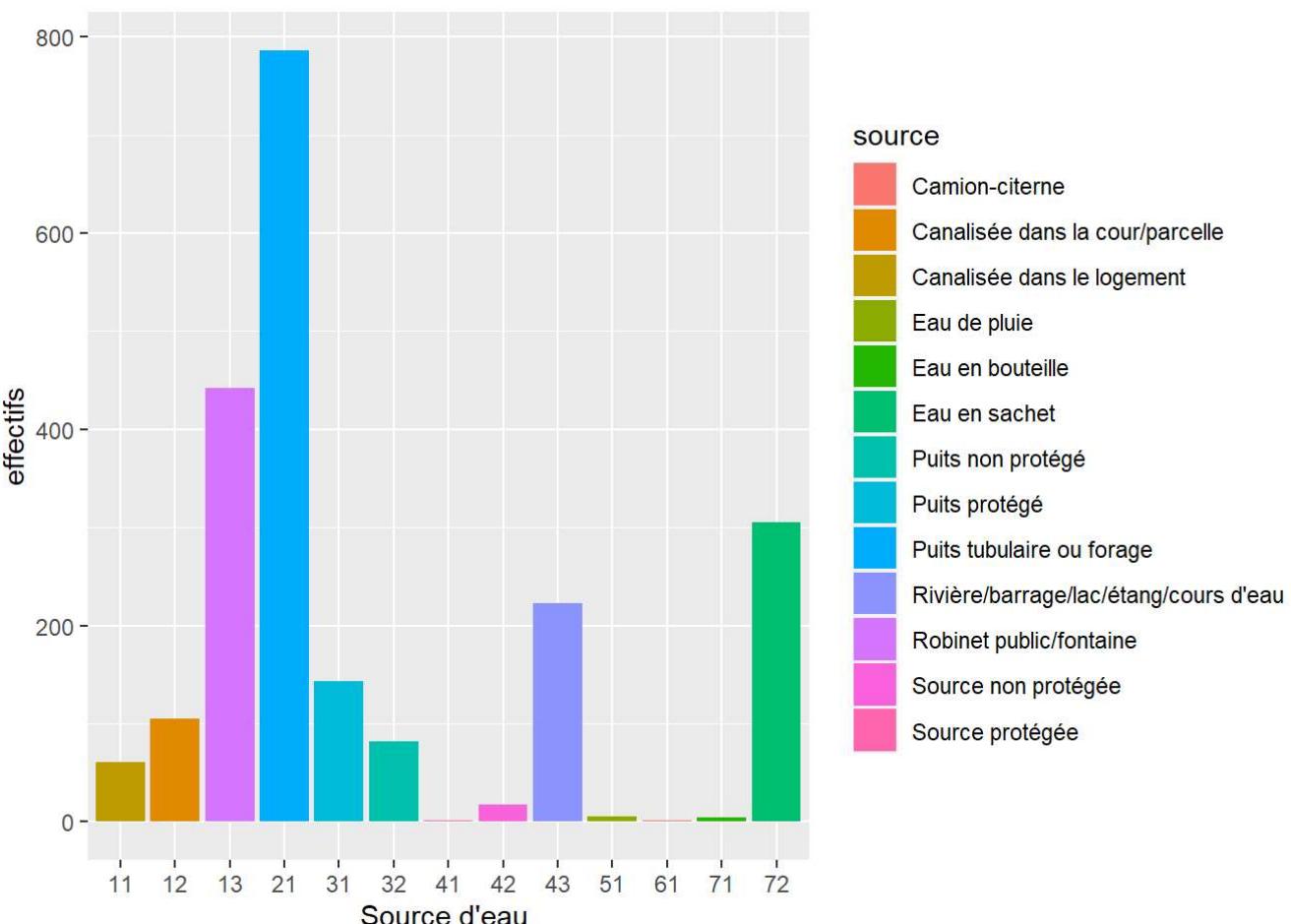
#Affichage du tableau des statistiques de la variable
var_source1

```

	## # A tibble: 13 × 6	## Source effectifs `effectifs cumulés` fréquences `fréquences cumulées` source	## <fct>	<int>	<dbl>
## 1	11	61	11	61	2.8
## 2	12	105	12	166	4.83
## 3	13	442	13	608	20.3
## 4	21	786	21	1394	36.1
## 5	31	143	31	1537	6.57
## 6	32	82	32	1619	3.77
## 7	41	1	41	1620	0.05
## 8	42	17	42	1637	0.78
## 9	43	223	43	1860	10.2
## 10	51	5	51	1865	0.23
## 11	61	1	61	1866	0.05
## 12	71	4	71	1870	0.18
## 13	72	305	72	2175	14.0

Diagramme

```
ggplot(var_source1,aes(x=Source, y=effectifs, fill=source))+ geom_bar(stat = "identity")+
  labs(x="Source d'eau ", y="effectifs")
```



D'après les données du tableau et du diagramme, il est possible de constater que l'eau potable est principalement prélevée à partir de puits ou de forages. En effet, 36,14 % des enfants appartiennent à des ménages interrogés qui tirent leur eau de ces sources. De plus, 20 % des enfants ont comme principale source d'eau les robinets publics et les fontaines, tandis que 10 % proviennent de ménages où l'eau potable est issue de rivières, lacs ou canaux. Seuls 7 % des enfants consomment de l'eau provenant d'une source ou d'un puit protégé.

Ainsi, on peut dire que la majorité de l'eau potable des ménages est issue de sources "naturelles" et nécessite d'être prélevée directement à la source. Cette situation indique un faible niveau d'hygiène de l'eau pour notre population, ce qui pourrait avoir des conséquences sur la santé.

Par la suite , lors du tris croisés et de l'ACM, nous allons utiliser le regroupement ci-dessus pour faciliter les interprétations.

```
# Regroupement des modalités
enf$source <- vector("character", length = nrow(enf))

enf$source[enf$cHV201==11]="Source canalisée"
enf$source[enf$cHV201==12]="Source canalisée"
enf$source[enf$cHV201==13]="Source canalisée"
enf$source[enf$cHV201==21]="Source protégée"
enf$source[enf$cHV201==31]="Source protégée"
enf$source[enf$cHV201==32]="Source non protégée"
enf$source[enf$cHV201==41]="Source protégée"
enf$source[enf$cHV201==42]="Source non protégée"
enf$source[enf$cHV201==43]="Source non protégée"
enf$source[enf$cHV201==51]="Source non protégée"
enf$source[enf$cHV201==61]="Eau en réservoir"
enf$source[enf$cHV201==71]="Eau en réservoir"
enf$source[enf$cHV201==72]="Eau en réservoir"

enf$cHV201 <- as.factor(enf$source)

# regroupement et calcul de l'effectif par modalité
var_source <- enf %>% group_by(cHV201)%>%
    summarize( effectifs=n())

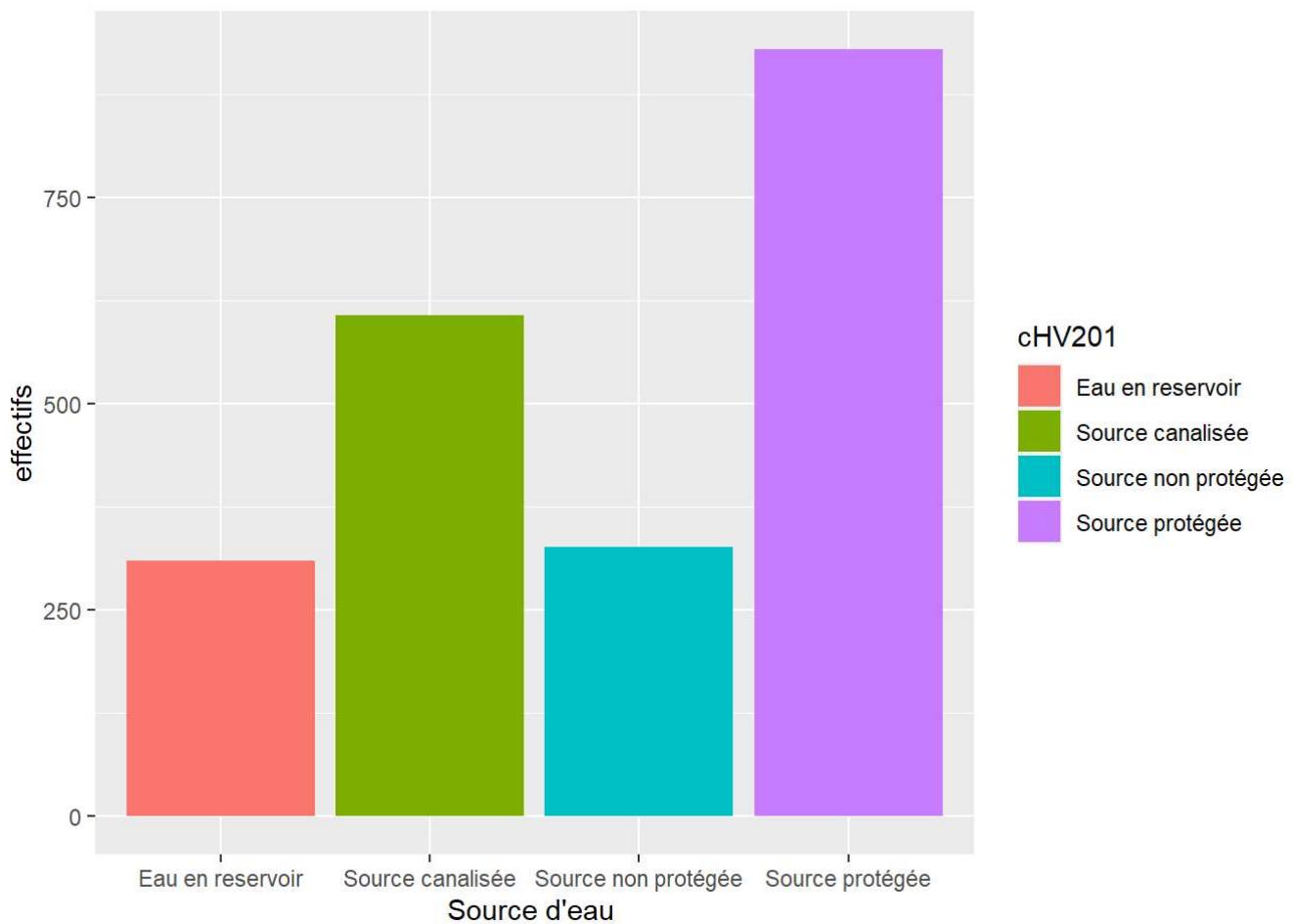
var_source$eff_cum <- cumsum(var_source$effectifs)

# Fréquences
var_source$pourcent <- round(prop.table(var_source$effectifs)*100,2)
var_source$pour_cum <- cumsum(var_source$pourcent)

# Affichage du tableau des statistiques de la variable
var_source
```

```
## # A tibble: 4 × 5
##   cHV201          effectifs  eff_cum pourcent pour_cum
##   <fct>            <int>     <int>    <dbl>     <dbl>
## 1 Eau en réservoir      310       310    14.2     14.2
## 2 Source canalisée      608       918    28.0     42.2
## 3 Source non protégée    327      1245    15.0     57.2
## 4 Source protégée       930      2175    42.8    100.
```

```
#Diagramme
ggplot(var_source,aes(x=cHV201, y=effectifs, fill=cHV201))+ geom_bar(stat = "identity")+
    labs(x="Source d'eau ", y="effectifs")
```



Etude de la variable Electricité

La variable “présence de l’électricité” est pertinente dans une étude sur l’anémie chez les enfants de moins de 5 ans. Elle offre des indices sur le niveau d’infrastructure, les conditions de vie, l’accès à l’éclairage et à la nutrition. La présence de l’électricité peut influencer la disponibilité d’appareils électroménagers pour la préparation et la conservation des aliments. Des conditions d’alimentation inadéquates liées au manque d’électricité peuvent contribuer aux carences nutritionnelles et à l’anémie. Ainsi, nous avons décidé de faire un tri à plat sur cette variable.

#electricité HV206

```

# Effectifs
var_elec <- enf %>% group_by(chV206)%>%
  summarize( effectifs=n())
var_elec$eff_cum <-cumsum(var_elec$effectifs)

# Fréquences
var_elec$pourcent <-round(prop.table(var_elec$effectifs)*100,2)
var_elec$pour_cum <- cumsum(var_elec$pourcent)

# Modalités
elec <- c("Non", "Oui")
var_elec$electricite <- elec

# Noms des colonnes
colnames(var_elec)<-c("Electricité", "effectifs","effectifs cumulés","fréquences","fréquences cumulées","electricite")

# Exportation de la table
write.xlsx(var_elec,file="var_elec.xlsx")

# Affichage du tableau des statistiques de la variable
var_elec

```

```

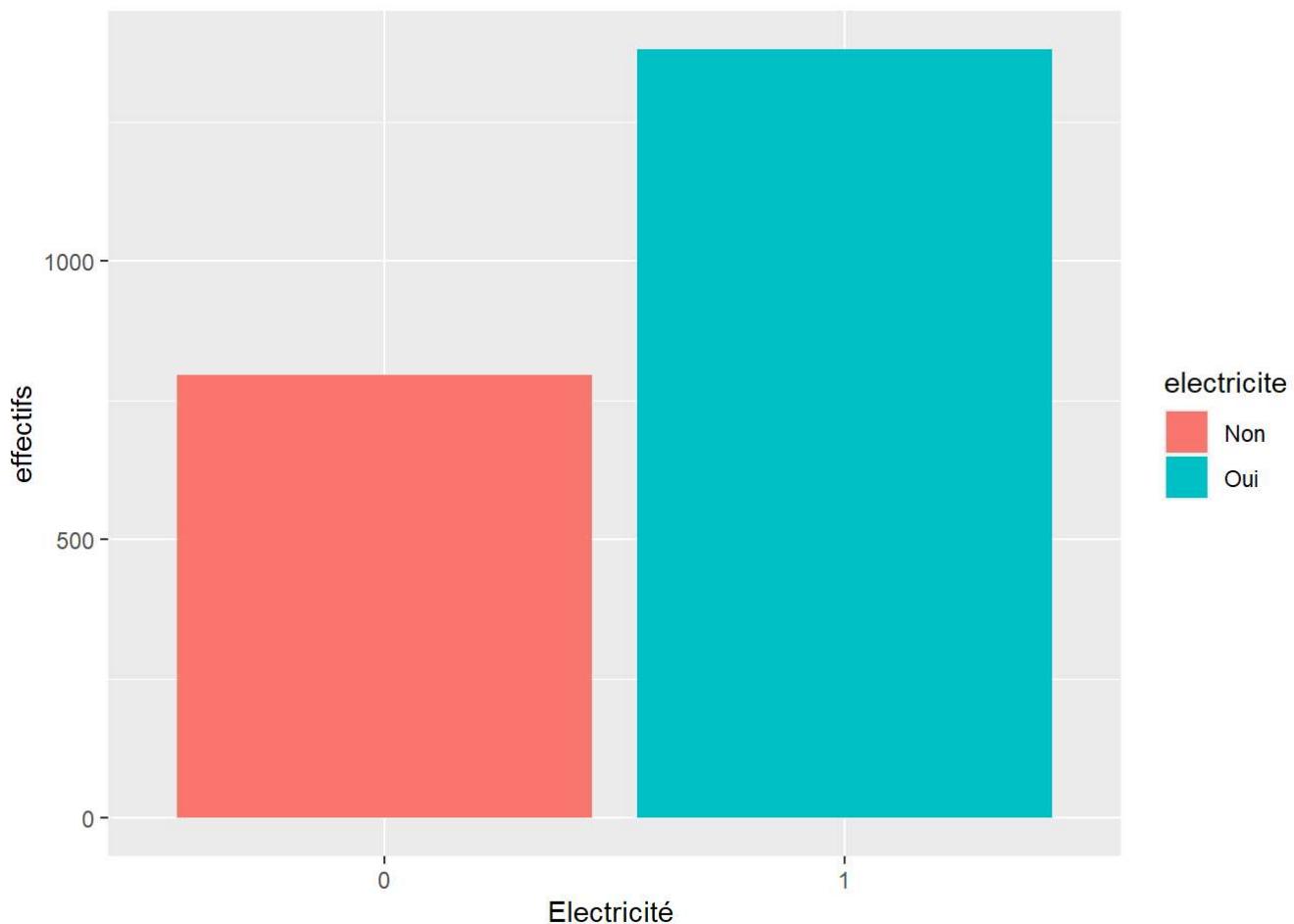
## # A tibble: 2 × 6
##   Electricité effectifs `effectifs cumulés` fréquences `fréquences cumulées`
##   <fct>      <int>           <int>      <dbl>           <dbl>
## 1 0            795             795       36.6            36.6
## 2 1           1380            2175      63.4            100
## # i 1 more variable: electricite <chr>

```

```

# Diagramme
ggplot(var_elec,aes(x=Electricité, y=effectifs, fill=electricite))+ geom_bar(stat = "identity")+
  labs(x="Electricité ", y="effectifs")

```



On remarque que 1380 enfants vivent dans des ménages équipés d'électricité, ce qui représente 63 % de notre population. En revanche, 795 enfants résident dans des logements sans électricité, soit 37 % de l'ensemble des enfants, un pourcentage non négligeable. Ainsi, un nombre considérable d'enfants vivent dans des ménages dépourvus d'électricité.

Etude de la variable Indice de richesse (HV270)

La richesse du ménage peut avoir un impact sur plusieurs déterminants de la santé des enfants, y compris leur alimentation, leur accès aux soins de santé, leurs conditions sanitaires, leur éducation sanitaire, et leur accès à une eau potable de qualité. Analyser cette variable est donc importante.

#Indice de richesse HV270

```

# Effectifs
var_rich <- enf %>%
  group_by(cHV270)%>%
  summarize( effectifs=n())
var_rich$eff_cum <-cumsum(var_rich$effectifs)

# Fréquences
var_rich$pourcent <-round(prop.table(var_rich$effectifs)*100,2)
var_rich$pour_cum <- cumsum(var_rich$pourcent)

# Modalités
richesse <- c("Très pauvre", "Pauvre","Modeste","Riche"," Plus riche" )
var_rich$richesse <- richesse

# Noms des colonnes
colnames(var_rich)<-c("Richesse", "effectifs","effectifs cumulés","fréquences","fréquences cumulées","richesse")

# Exportation de la table
write.xlsx(var_rich,file="var_rich.xlsx")

# Affichage de la table
var_rich

```

```

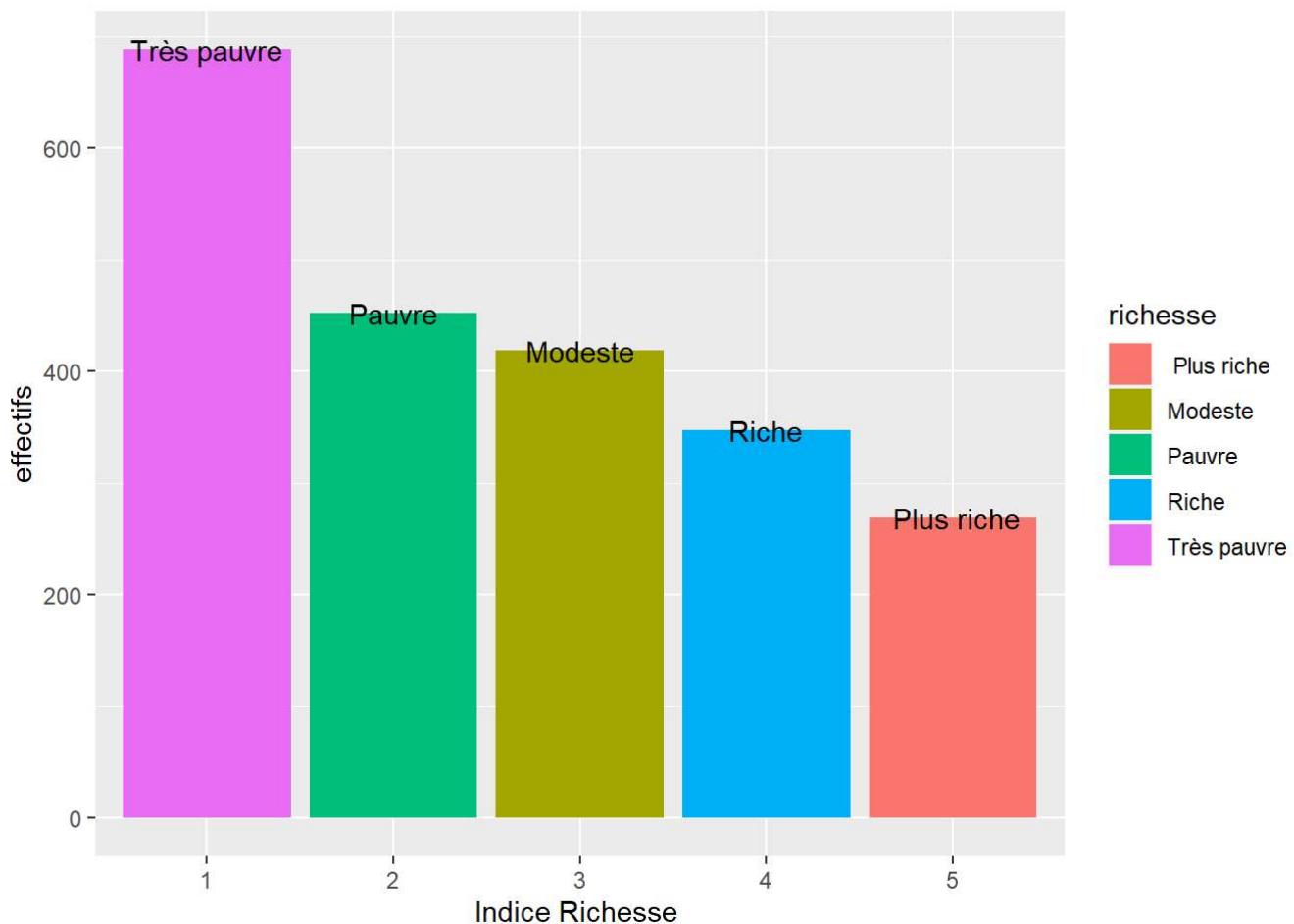
## # A tibble: 5 × 6
##   Richesse effectifs `effectifs cumulés` fréquences `fréquences cumulées`
##   <fct>     <int>             <int>      <dbl>            <dbl>
## 1 1           688              688       31.6            31.6
## 2 2           452             1140      20.8            52.4
## 3 3           419             1559      19.3            71.7
## 4 4           347             1906      16.0            87.6
## 5 5           269             2175      12.4            100.
## # i 1 more variable: richesse <chr>

```

```

# Diagramme
ggplot(var_rich,aes(x=Richesse, y=effectifs, fill=richesse))+ geom_bar(stat = "identity") +
  labs(x="Indice Richesse ", y="effectifs")+ geom_text(label=richesse)

```



D'après le diagramme et le tableau, on observe que plus de la moitié des enfants de moins de 5 ans sont issues des menages pauvres. Environ 19 % des enfants vivent dans des menages de niveau de richesse modeste et 28% vivent dans des menages riches. On a donc que la majorité des enfants (71,67%) vivent dans des menages à revenus faibles ou modestes .

Etude de la variable Anémie

L'étude de la variable anémie dans ce travail revêt une grande importance car l'anémie peut être un indicateur significatif de la santé nutritionnelle des enfants. L'analyse de cette variable va nous offrir des informations cruciales sur la santé nutritionnelle, les habitudes alimentaires et le bien-être global des enfants. Il s'agit donc de la variable la plus importante de notre étude, c'est la variable que l'on va chercher à expliquer.

```

# Effectifs
var_anemie <- enf %>% group_by(cHA57)%>%
    summarize( effectifs=n())

var_anemie$eff_cum <-cumsum(var_anemie$effectifs)

# Fréquences
var_anemie$pourcent <-round(prop.table(var_anemie$effectifs)*100,2)
var_anemie$pour_cum <- cumsum(var_anemie$pourcent)

# Modalités
niv <- c("Sevère", "Moderé", "Leger", "Non anémique" )
var_anemie$niv_anemie <- niv

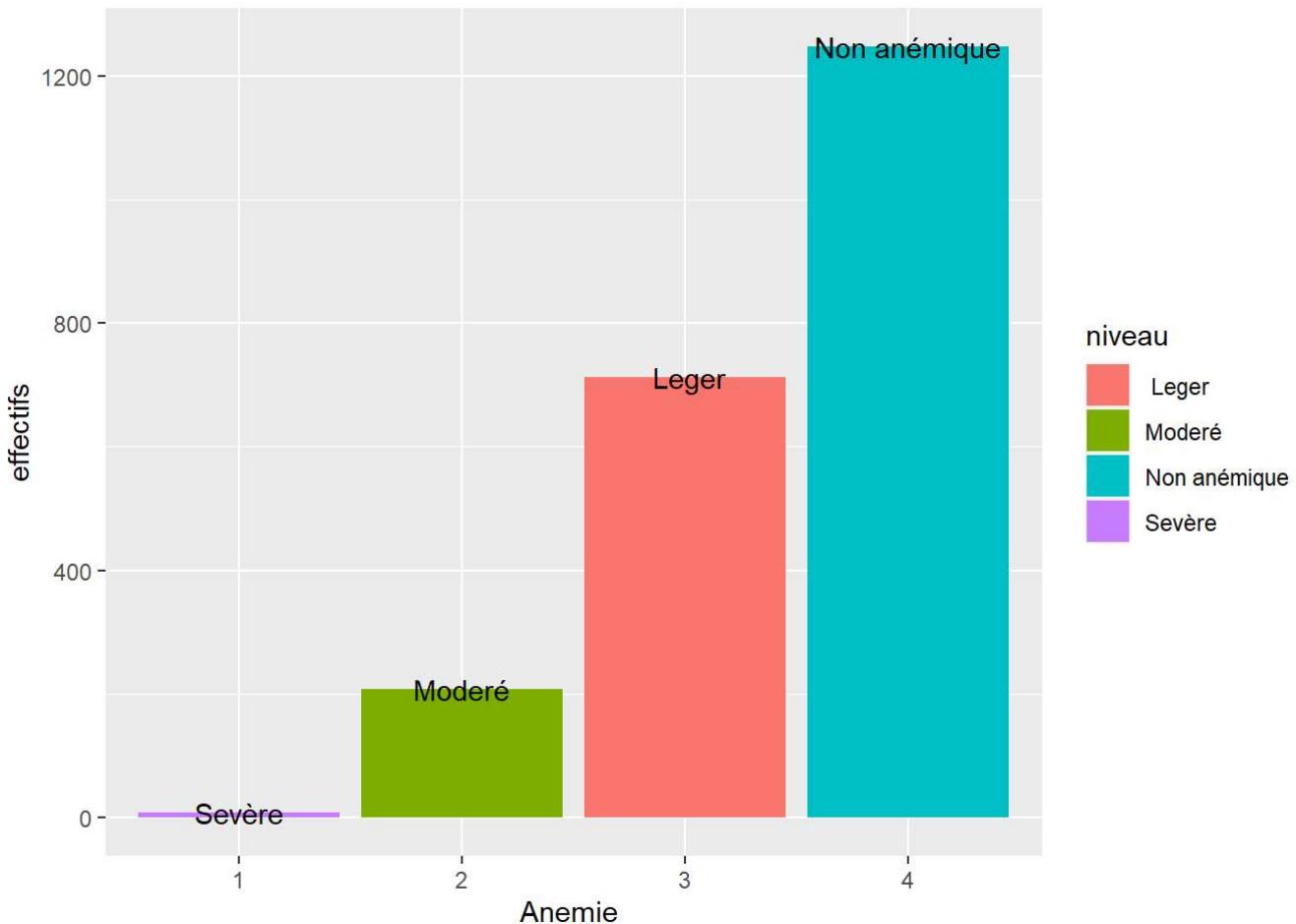
# Noms des colonnes
colnames(var_anemie)<-c("Anémie", "effectifs","effectifs cumulés","fréquences","fréquences cumulées","niveau")

# Exportation de la table
write.xlsx(var_anemie,file="var_anemie.xlsx")

# Affichage du tableau des statistiques de la variable:

# Diagramme
ggplot(var_anemie,aes(x=Anémie, y=effectifs, fill=niveau))+ geom_bar(stat = "identity") +
  labs(x="Anémie ", y="effectifs")+ geom_text(label=niv)

```



D'après avoir observer l'histogramme et le tableau ci-dessus, on remarque que, 57 % des enfants de moins de 5ans sont non anémiques. Les enfants qui ont une anémie de niveau modéré ou sevère représentent 10% et

32% des enfants souffrent d'une anémie légère. On a au total que 42% des enfants souffrent d'anémie.

On peut affirmer que le taux d'enfants souffrant d'anémie est significatif. Par la suite, nous effectuerons des analyses croisées de la variable "anémie" avec d'autres variables pour déterminer s'il existe des liens entre cette dernière et les autres variables.

Tris croisés :

Pour faciliter , l'étude et pouvoir faire le test de khi-2, nous avons décidé de regrouper certaines modalités de la variable Anémie qui présente des faibles effectifs.

```
summary(enf$cHA57)
```

```
##    1    2    3    4
##    8  208  712 1247
```

```
enf$HA57[enf$HA57==1]="Modéré"
enf$HA57[enf$HA57==2 ]="Modéré"
enf$HA57[enf$HA57==3 ]="Leger"
enf$HA57[enf$HA57==4 ]="Non anémique"
enf$cHA57 <-as.factor(enf$HA57)
```

Croisement des variables anémie et niveau de richesse du ménage

Nous pensons qu'il pourrait y avoir une corrélation entre l'indice de richesse du ménage et l'anémie chez l'enfant. En examinant ces deux variables simultanément, nous pourrons déterminer s'il existe une relation entre elles.

```
library(gmodels)
```

```
## Warning: le package 'gmodels' a été compilé avec la version R 4.3.2
```

```
anemie_rich <- table(enf$cHA57,enf$cHV270)

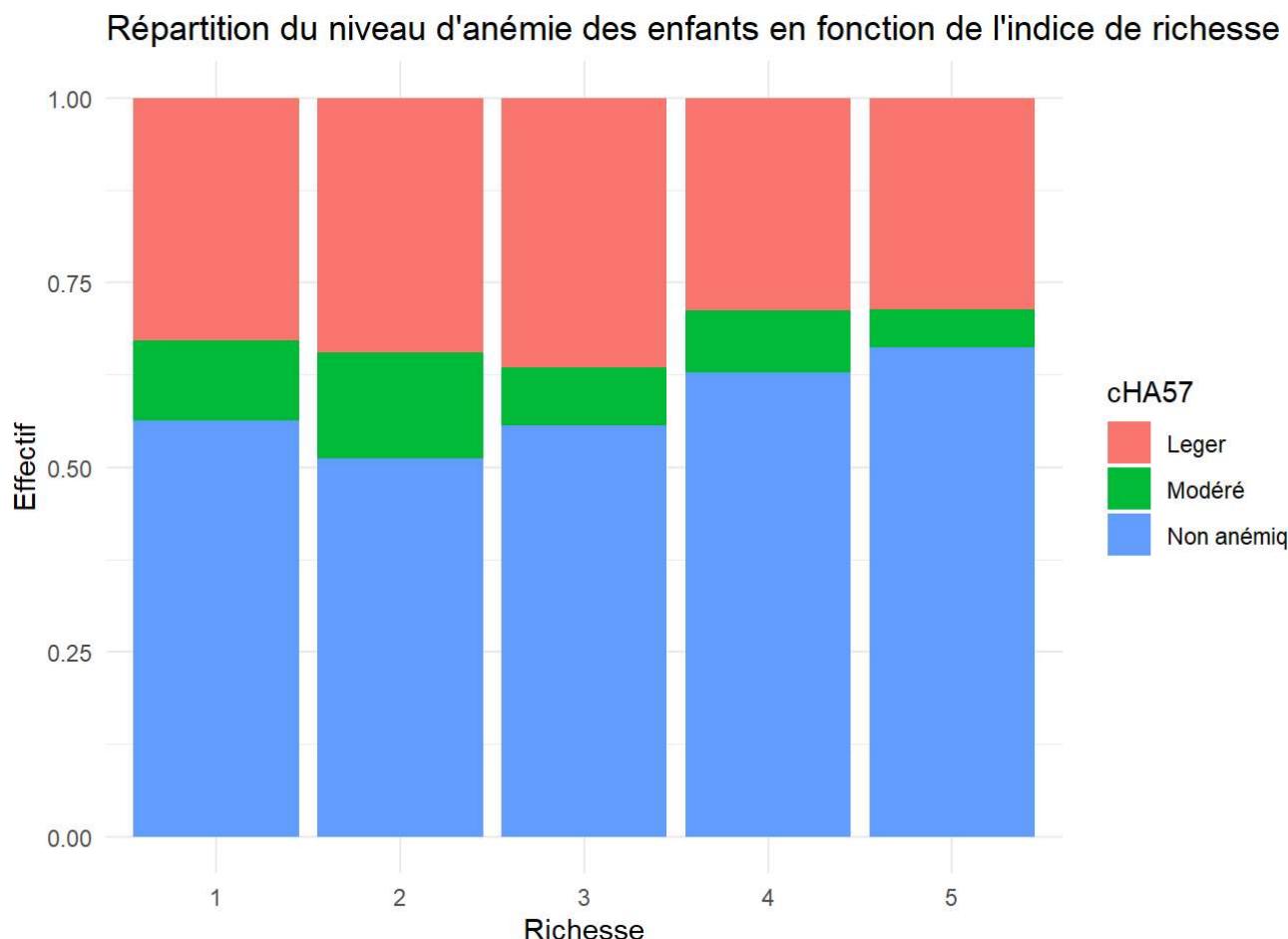
enf$anemie <- vector("character", length = nrow(enf))
enf$anemie[enf$cHA57=="Modéré"]="Anémique"
enf$anemie[enf$cHA57=="Modéré "]="Anémique"
enf$anemie[enf$cHA57=="Leger "]="Anémique"
enf$anemie[enf$cHA57=="Non anémique"]="Non anémique"

enf$anemie <-as.factor(enf$anemie)
```

```
prop.table(table(enf$cHA57,enf$cHV270))
```

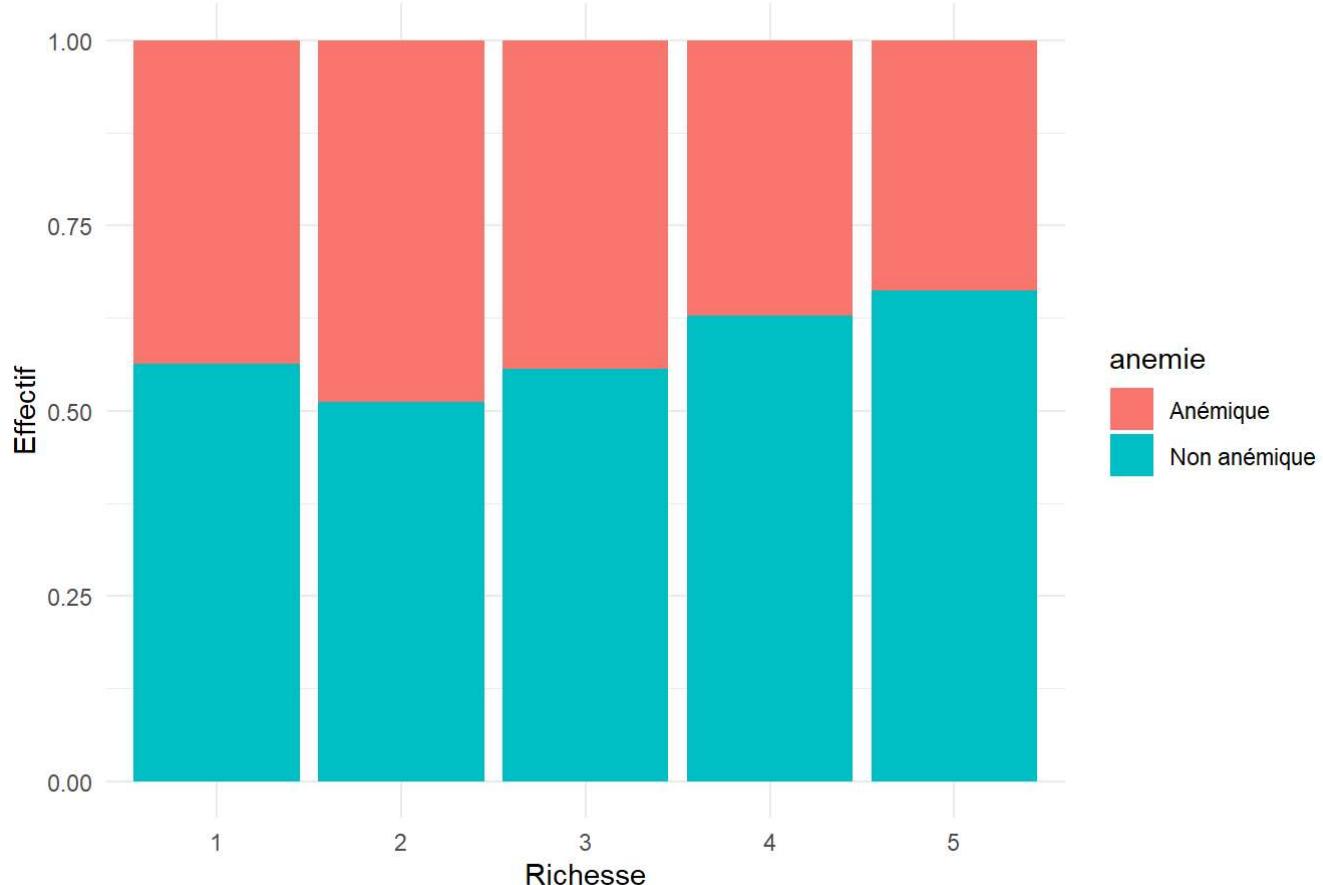
```
##  
##  
##  
##   Leger      0.103908046 0.071724138 0.070344828 0.045977011 0.035402299  
##   Modéré     0.034482759 0.029885057 0.015172414 0.013333333 0.006436782  
##   Non anémique 0.177931034 0.106206897 0.107126437 0.100229885 0.081839080
```

```
ggplot(enf) +  
  aes(x = cHV270, fill = cHA57) +  
  geom_bar(position = "fill") +  
  scale_fill_hue(direction = 1) +  
  labs(x = "Richesse", y = "Effectif") +  
  theme_minimal() +  
  ggtitle("Répartition du niveau d'anémie des enfants en fonction de l'indice de richesse")
```



```
ggplot(enf) +  
  aes(x = cHV270, fill = anemie) +  
  geom_bar(position = "fill") +  
  scale_fill_hue(direction = 1) +  
  labs(x = "Richesse", y = "Effectif") +  
  theme_minimal() +  
  ggtitle("Répartition d'anémie des enfants en fonction de l'indice de richesse")
```

Répartition d'anémie des enfants en fonction de l'indice de richesse



On remarque en visualisant cette figure de répartition du niveau d'anémie en fonction de la richesse du ménage que le pourcentage des enfants anémiques est importante dans les ménages pauvres ou modestes que dans les menages riches. Et pour vérifier la liaison entre ces deux facteurs, nous réalisons un test de Khi-deux d'indépendance en posant deux hypothèses :

Hypothèse nulle : les deux variables sont indépendantes Hypothèse alternative : les deux variables ne sont pas indépendantes

#test khi-2 Niveau d'anémie Niveau de richesse

```
CrossTable(enf$cHA57,enf$cHV270,chisq = T, dnn = c("Niveau d'anémie","Niveau de richesse"))
```

```

## 
## 
##   Cell Contents
## |-----|
## |           N |
## | Chi-square contribution |
## |           N / Row Total |
## |           N / Col Total |
## |           N / Table Total |
## |-----|
## 
## 
## Total Observations in Table: 2175
## 
## 
##          | Niveau de richesse
## Niveau d'anémie |    1 |    2 |    3 |    4 |    5 | Row Total |
## -----|-----|-----|-----|-----|-----|-----|
##     Leger | 226 | 156 | 153 | 100 | 77 | 712 |
##             | 0.003 | 0.436 | 1.829 | 1.627 | 1.389 | 
##             | 0.317 | 0.219 | 0.215 | 0.140 | 0.108 | 0.327 |
##             | 0.328 | 0.345 | 0.365 | 0.288 | 0.286 | 
##             | 0.104 | 0.072 | 0.070 | 0.046 | 0.035 | 
## -----|-----|-----|-----|-----|-----|-----|
##     Modéré | 75 | 65 | 33 | 29 | 14 | 216 |
##             | 0.652 | 9.011 | 1.782 | 0.865 | 6.051 | 
##             | 0.347 | 0.301 | 0.153 | 0.134 | 0.065 | 0.099 |
##             | 0.109 | 0.144 | 0.079 | 0.084 | 0.052 | 
##             | 0.034 | 0.030 | 0.015 | 0.013 | 0.006 | 
## -----|-----|-----|-----|-----|-----|-----|
## Non anémique | 387 | 231 | 233 | 218 | 178 | 1247 |
##             | 0.141 | 3.057 | 0.217 | 1.825 | 3.665 | 
##             | 0.310 | 0.185 | 0.187 | 0.175 | 0.143 | 0.573 |
##             | 0.562 | 0.511 | 0.556 | 0.628 | 0.662 | 
##             | 0.178 | 0.106 | 0.107 | 0.100 | 0.082 | 
## -----|-----|-----|-----|-----|-----|-----|
## Column Total | 688 | 452 | 419 | 347 | 269 | 2175 |
##             | 0.316 | 0.208 | 0.193 | 0.160 | 0.124 | 
## -----|-----|-----|-----|-----|-----|-----|
## 
## 
## Statistics for All Table Factors
## 
## 
## Pearson's Chi-squared test
## 
## -----|
## Chi^2 = 32.54918      d.f. = 8      p = 7.423668e-05
## 
## 
## 
## 
```

Sur tableau du test du Khi-deux, nous observons que la p-Valeur du test est de 0.06% ce qui est inférieur à 5%, nous pouvons donc rejeter l'hypothèse nulle. Il existe donc un lien entre le niveau de richesse du ménage et l'anémie des enfants.

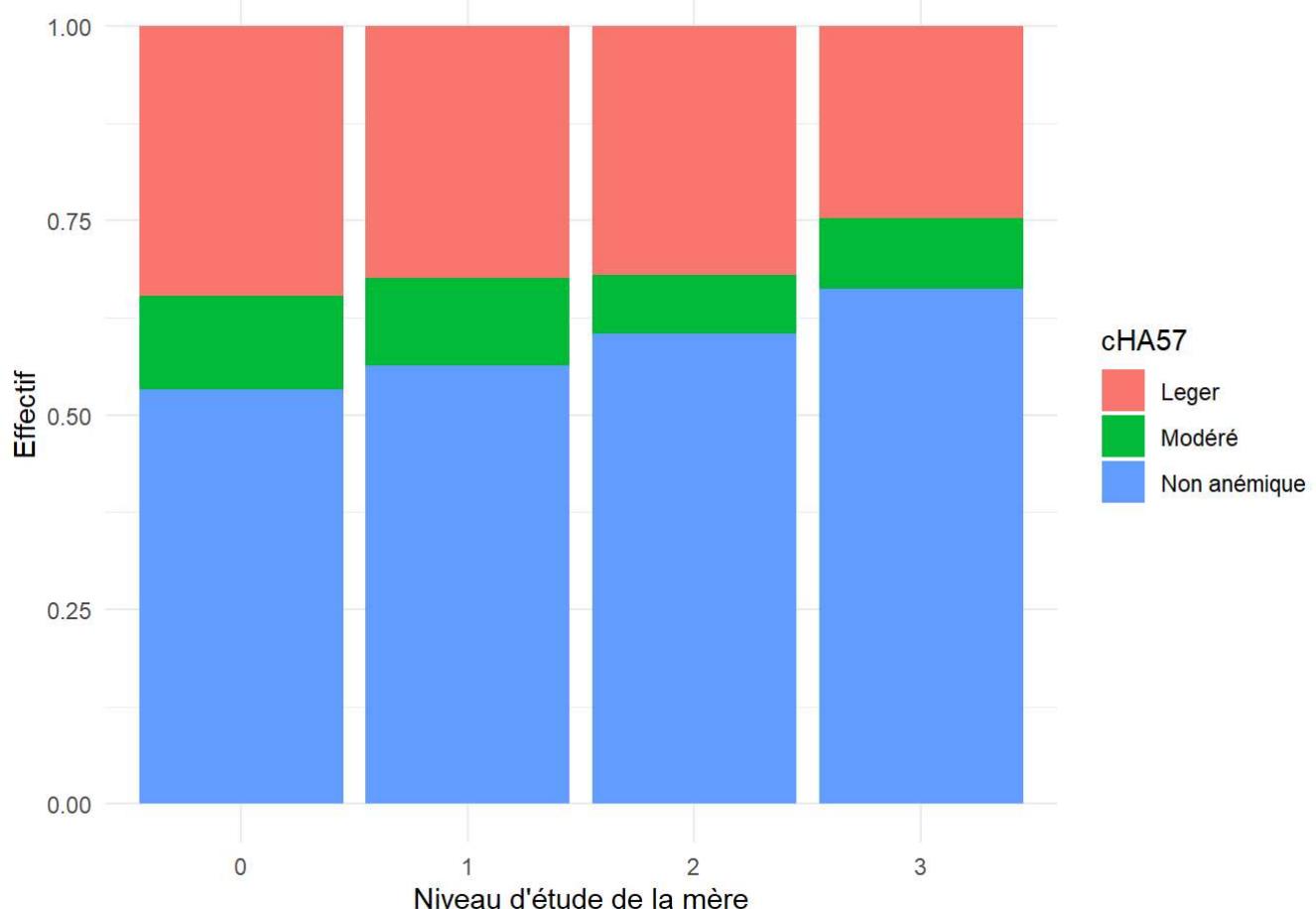
Croisement anémie et niveau d'étude de la mère

Cette variable est intéressante de notre étude, car les mères ayant un niveau d'éducation plus élevé sont souvent mieux informées sur les besoins nutritionnels de leurs enfants. Elles sont susceptibles d'adopter des pratiques alimentaires plus équilibrées, ce qui contribue à prévenir l'anémie chez les enfants. C'est pourquoi nous avons décidé d'explorer la corrélation entre l'anémie et le niveau d'éducation des mères.

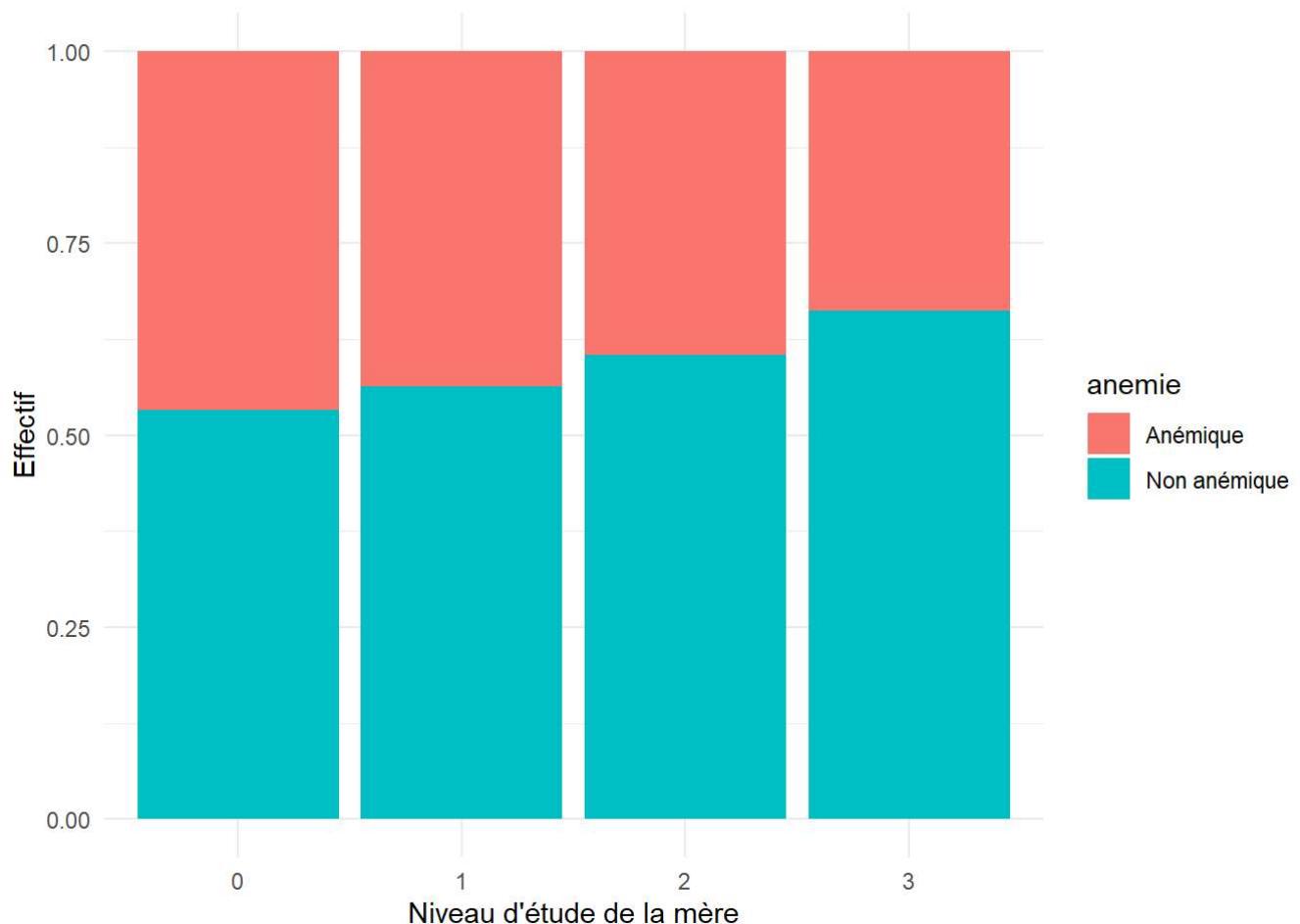
```
anemie_etud <- table(enf$cHA57,enf$cHC61)
table(enf$cHA57,enf$cHC61)
```

```
##          0   1   2   3
## Leger     264 141 288 19
## Modéré    91  49  69  7
## Non anémique 406 245 545 51
```

```
ggplot(enf) +
  aes(x = cHC61, fill = cHA57) +
  geom_bar(position = "fill") +
  scale_fill_hue(direction = 1) +
  labs(x = "Niveau d'étude de la mère",y="Effectif")+
  theme_minimal()
```



```
ggplot(enf) +
  aes(x = cHC61, fill = anemie) +
  geom_bar(position = "fill") +
  scale_fill_hue(direction = 1) +
  labs(x = "Niveau d'étude de la mère", y = "Effectif") +
  theme_minimal()
```



```
CrossTable(enf$cHA57, enf$cHC61, chisq = T, dnn = c("Niveau d'anémie", "Niveau d'étude de la mère"))
```

```

## 
## 
##   Cell Contents
## |-----|
## |           N |
## | Chi-square contribution |
## |           N / Row Total |
## |           N / Col Total |
## |           N / Table Total |
## |-----|
## 
## 
## Total Observations in Table: 2175
## 
## 
##          | Niveau d'étude de la mère
## Niveau d'anémie |     0 |     1 |     2 |     3 | Row Total |
## -----|-----|-----|-----|-----|-----|
##       Leger | 264 | 141 | 288 | 19 | 712 |
##             | 0.889 | 0.014 | 0.179 | 1.528 | 
##             | 0.371 | 0.198 | 0.404 | 0.027 | 0.327 |
##             | 0.347 | 0.324 | 0.319 | 0.247 | 
##             | 0.121 | 0.065 | 0.132 | 0.009 | 
## -----|-----|-----|-----|-----|-----|
##       Modéré | 91 | 49 | 69 | 7 | 216 |
##             | 3.148 | 0.779 | 4.727 | 0.055 | 
##             | 0.421 | 0.227 | 0.319 | 0.032 | 0.099 |
##             | 0.120 | 0.113 | 0.076 | 0.091 | 
##             | 0.042 | 0.023 | 0.032 | 0.003 | 
## -----|-----|-----|-----|-----|-----|
## Non anémique | 406 | 245 | 545 | 51 | 1247 |
##             | 2.105 | 0.078 | 1.500 | 1.064 | 
##             | 0.326 | 0.196 | 0.437 | 0.041 | 0.573 |
##             | 0.534 | 0.563 | 0.604 | 0.662 | 
##             | 0.187 | 0.113 | 0.251 | 0.023 | 
## -----|-----|-----|-----|-----|-----|
## Column Total | 761 | 435 | 902 | 77 | 2175 |
##             | 0.350 | 0.200 | 0.415 | 0.035 | 
## -----|-----|-----|-----|-----|-----|
## 
## 
## Statistics for All Table Factors
## 
## 
## Pearson's Chi-squared test
## 
## -----|
## Chi^2 = 16.06588      d.f. = 6      p = 0.01340467
## 
## 
## 
## 
```

À partir de la figure ci-dessus, il est observé que plus le niveau d'éducation de la mère est élevé, moins le niveau d'anémie de l'enfant semble être élevé. Par ailleurs, on observe un taux significatif de cas d'anémie modérée ou sévère chez les enfants issus de mères ayant un niveau éducatif bas. Et pour vérifier la liaison entre ces deux facteurs, nous réalisons un test de Khi-deux d'indépendance en posant deux hypothèses :

Hypothèse nulle : les deux variables sont indépendantes Hypothèse alternative : les deux variables ne sont pas indépendantes

#khi-2 anemie_etud

```
CrossTable(enf$cHA57,enf$cHC61,chisq = T, dnn = c("Niveau d'anémie","Niveau d'étude de la mère"))
```

```

## 
## 
##   Cell Contents
## |-----|
## |           N |
## | Chi-square contribution |
## |           N / Row Total |
## |           N / Col Total |
## |           N / Table Total |
## |-----|
## 
## 
## Total Observations in Table: 2175
## 
## 
##          | Niveau d'étude de la mère
## Niveau d'anémie |     0 |     1 |     2 |     3 | Row Total |
## -----|-----|-----|-----|-----|-----|
##       Leger | 264 | 141 | 288 | 19 | 712 |
##             | 0.889 | 0.014 | 0.179 | 1.528 | 
##             | 0.371 | 0.198 | 0.404 | 0.027 | 0.327 |
##             | 0.347 | 0.324 | 0.319 | 0.247 | 
##             | 0.121 | 0.065 | 0.132 | 0.009 | 
## -----|-----|-----|-----|-----|-----|
##       Modéré | 91 | 49 | 69 | 7 | 216 |
##             | 3.148 | 0.779 | 4.727 | 0.055 | 
##             | 0.421 | 0.227 | 0.319 | 0.032 | 0.099 |
##             | 0.120 | 0.113 | 0.076 | 0.091 | 
##             | 0.042 | 0.023 | 0.032 | 0.003 | 
## -----|-----|-----|-----|-----|-----|
## Non anémique | 406 | 245 | 545 | 51 | 1247 |
##             | 2.105 | 0.078 | 1.500 | 1.064 | 
##             | 0.326 | 0.196 | 0.437 | 0.041 | 0.573 |
##             | 0.534 | 0.563 | 0.604 | 0.662 | 
##             | 0.187 | 0.113 | 0.251 | 0.023 | 
## -----|-----|-----|-----|-----|-----|
## Column Total | 761 | 435 | 902 | 77 | 2175 |
##             | 0.350 | 0.200 | 0.415 | 0.035 | 
## -----|-----|-----|-----|-----|-----|
## 
## 
## Statistics for All Table Factors
## 
## 
## Pearson's Chi-squared test
## 
## -----|
## Chi^2 = 16.06588      d.f. = 6      p = 0.01340467
## 
## 
## 
## 
```

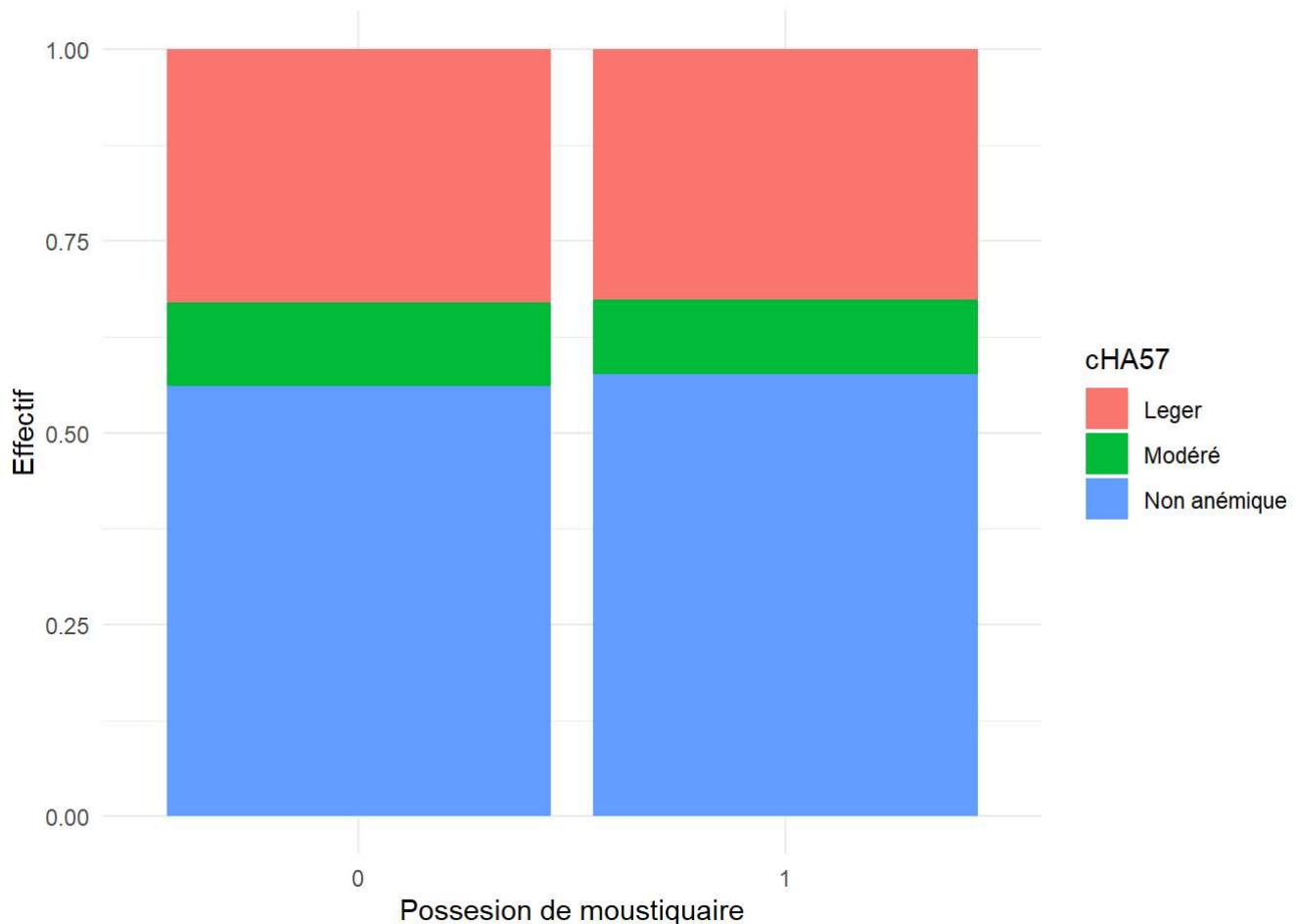
En examinant le tableau du test du Chi-deux, on constate que la p-valeur est de 1,3%, ce qui est inférieur à 5%. Par conséquent, nous pouvons rejeter l'hypothèse nulle, indiquant ainsi l'existence d'une relation significative entre le niveau d'éducation de la mère et l'anémie chez les enfants.

Croisement des variables anémie et possession de moustiquaire

Plusieurs études scientifiques ont démontré l'existence d'une relation entre le paludisme et l'anémie, soulignant que la présence du paludisme chez un individu aggrave ou provoque l'anémie. Initialement, notre intention était d'analyser la variable du paludisme. Cependant, en constatant l'existence de 25 colonnes (HML32_1, HML32_2, ..., HML32_3), nous n'avons pas été en mesure de distinguer les colonnes correspondant aux enfants. Étant donné que le paludisme est principalement transmis par les moustiques, nous avons donc décidé d'examiner si l'utilisation de moustiquaires pouvait réduire l'état anémique chez les enfants.

```
enf$HV228[is.na(enf$HV228)] <- 0
anemie_moustiq<- table(enf$cHA57,enf$cHV227)

ggplot(enf) +
  aes(x = cHV227, fill = cHA57) +
  geom_bar(position="fill") +
  scale_fill_hue(direction = 1) +
  labs(x = "Possession de moustiquaire",y="Effectif")+
  theme_minimal()
```



```
CrossTable(enf$cHA57,enf$HV227,chisq = T, dnn = c("Niveau d'anémie","Possession de moustiquaire"))
```

```

## 
## 
##   Cell Contents
## |-----|
## |           N |
## | Chi-square contribution |
## |           N / Row Total |
## |           N / Col Total |
## |           N / Table Total |
## |-----|
## 
## 
## Total Observations in Table:  2175
## 
## 
##          Possession de moustiquaire
## Niveau d'anémie |     0 |      1 | Row Total |
## -----|-----|-----|-----|
##       Leger | 134 | 578 | 712 |
##             | 0.009 | 0.002 | |
##             | 0.188 | 0.812 | 0.327 |
##             | 0.330 | 0.327 | |
##             | 0.062 | 0.266 | |
## -----|-----|-----|-----|
##       Modéré | 44 | 172 | 216 |
##             | 0.336 | 0.077 | |
##             | 0.204 | 0.796 | 0.099 |
##             | 0.108 | 0.097 | |
##             | 0.020 | 0.079 | |
## -----|-----|-----|-----|
## Non anémique | 228 | 1019 | 1247 |
##             | 0.098 | 0.022 | |
##             | 0.183 | 0.817 | 0.573 |
##             | 0.562 | 0.576 | |
##             | 0.105 | 0.469 | |
## -----|-----|-----|-----|
## Column Total | 406 | 1769 | 2175 |
##             | 0.187 | 0.813 | |
## -----|-----|-----|-----|
## 
## 
## Statistics for All Table Factors
## 
## 
## Pearson's Chi-squared test
## 
## -----
```

Chi^2 = 0.5443658 d.f. = 2 p = 0.7617149

##

##

##

En observant la figure ci-dessus, aucune disparité de pourcentage d'enfants anémiques n'est apparente entre les ménages équipés de moustiquaires et ceux qui n'en possèdent pas. De plus, selon le test du Khi-deux d'indépendance entre ces deux variables, la p-valeur est de 76%. Ainsi, nous ne pouvons pas rejeter l'hypothèse d'indépendance entre le niveau d'anémie et la possession de moustiquaires.

Croisement des variables anémie et région

Comme nous l'avons déjà souligné, le niveau d'anémie chez les enfants peut également être lié aux facteurs environnementaux et socio-économiques de chaque région.

```
anemie_region <-table(enf$anemie,enf$cHV024)
anemie_region
```

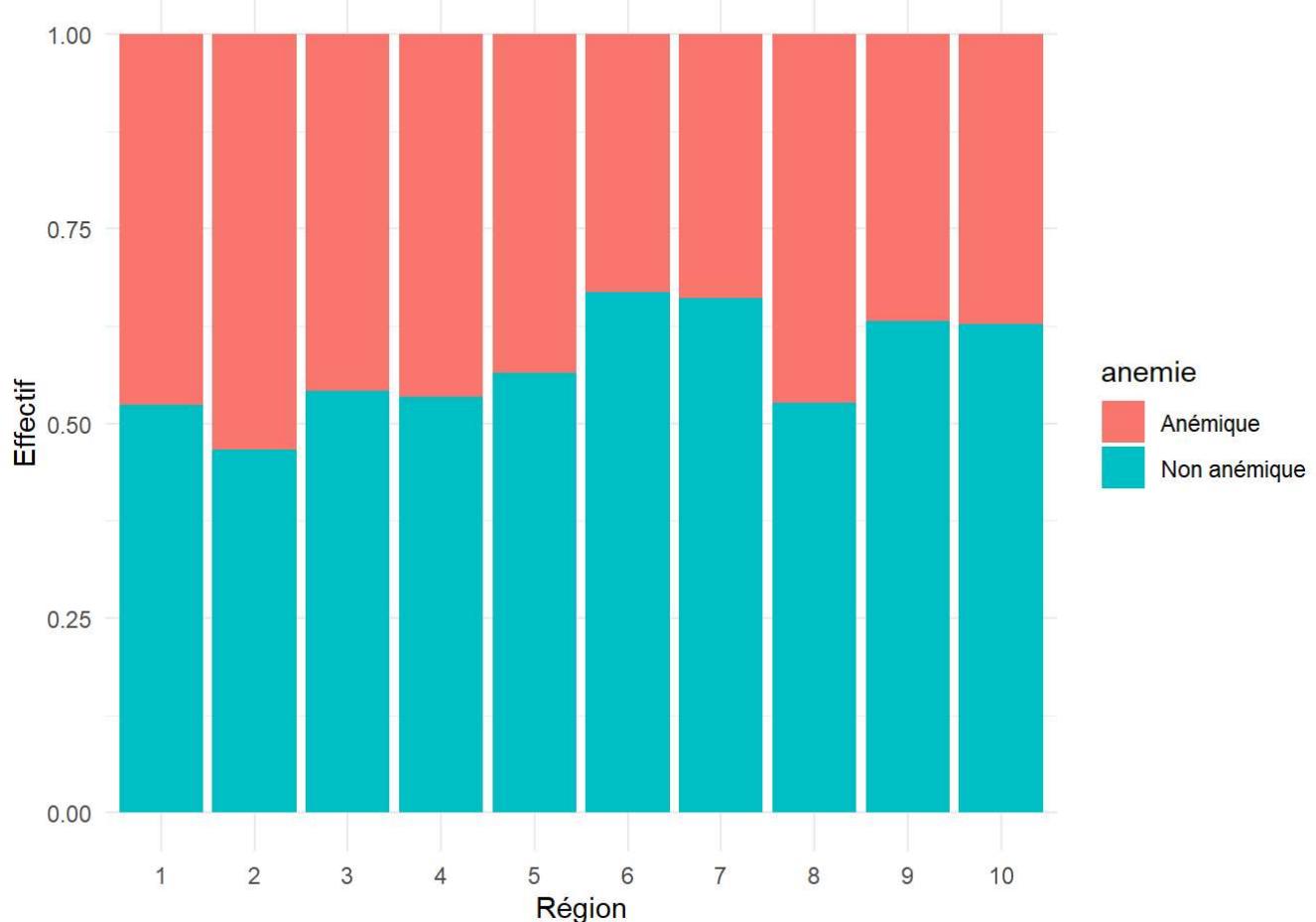
```
##
##          1   2   3   4   5   6   7   8   9   10
## Anémique    111 119  83  76  77  70  86 159  77  70
## Non anémique 122 104  98  87 100 141 168 177 132 118
```

```
tab_reg <-data.frame(ID=c(1:10), region=c("Western","Central","Greater Accra ", "Volta ","Eastern ","Ashanti ","Brong Ahafo","Northern","Upper East","Upper West"))

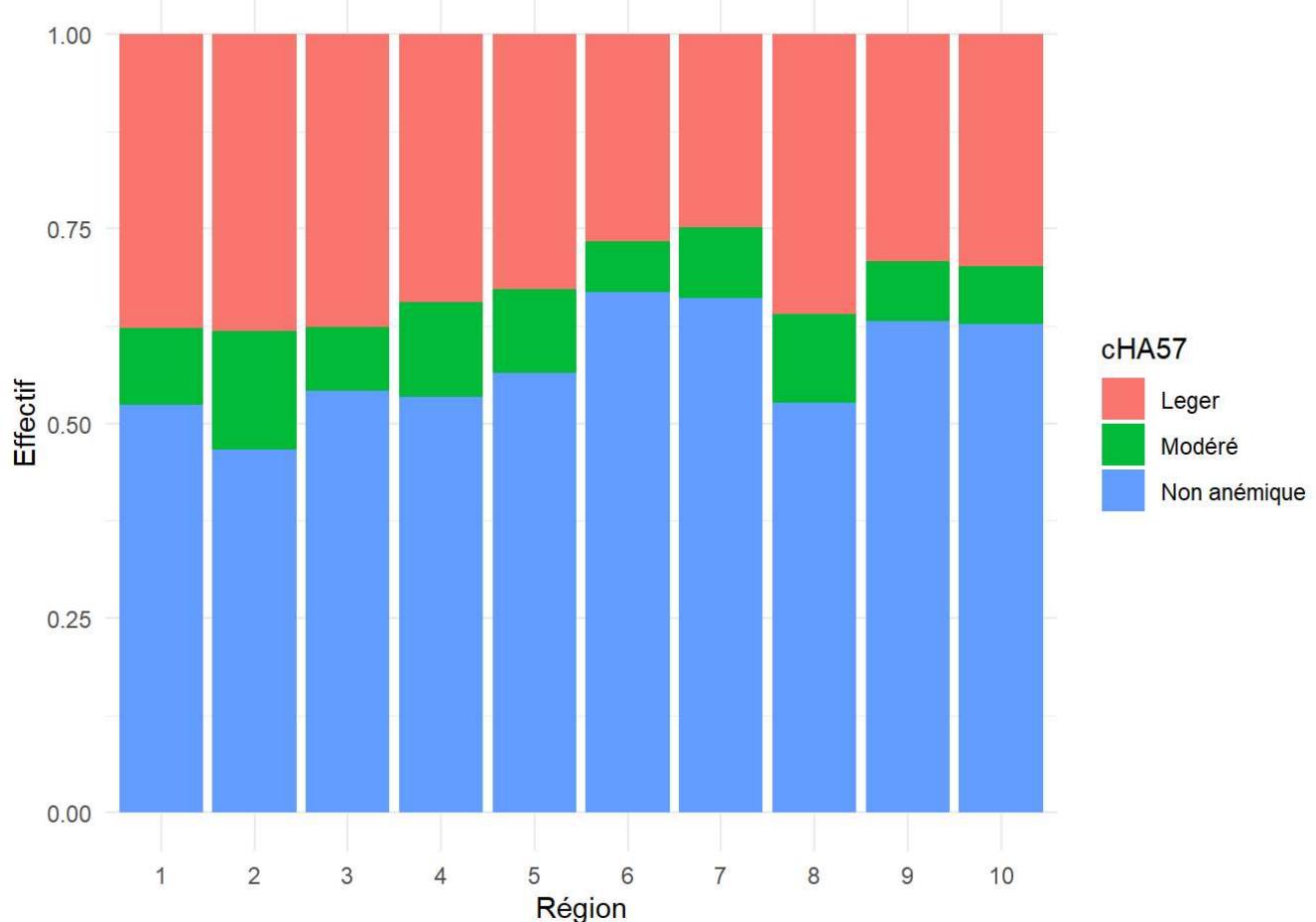
tab_reg
```

```
##      ID      region
## 1     1      Western
## 2     2      Central
## 3     3 Greater Accra
## 4     4      Volta
## 5     5      Eastern
## 6     6      Ashanti
## 7     7 Brong Ahafo
## 8     8      Northern
## 9     9      Upper East
## 10   10      Upper West
```

```
ggplot(enf) +
  aes(x = cHV024, fill = anemie) +
  geom_bar(position = "fill") +
  scale_fill_hue(direction = 1) +
  labs(x = "Région",y="Effectif")+
  theme_minimal()
```



```
ggplot(enf) +
  aes(x = cHV024, fill = cHA57) +
  geom_bar(position = "fill") +
  scale_fill_hue(direction = 1) +
  labs(x = "Région", y = "Effectif") +
  theme_minimal()
```



```
CrossTable(enf$cHA57,enf$cHV024, chisq = T, dnn = c("Niveau d'anémie","region"))
```

```

##  

##  

##      Cell Contents  

## |-----|  

## |           N |  

## | Chi-square contribution |  

## |           N / Row Total |  

## |           N / Col Total |  

## |           N / Table Total |  

## |-----|  

##  

##  

## Total Observations in Table: 2175  

##  

##  

##          | region  

## Niveau d'anémie |   1 |   2 |   3 |   4 |   5 |   6 |  

7 |     8 |    9 |   10 | Row Total |  

## -----|-----|-----|-----|-----|-----|-----|  

##          Leger |    88 |    85 |    68 |    56 |    58 |    56 |  

63 |   121 |   61 |   56 |   712 |  

##          | 1.803 | 1.972 | 1.292 | 0.131 | 0.000 | 2.474 |  

4.882 | 1.102 | 0.804 | 0.499 |  

##          | 0.124 | 0.119 | 0.096 | 0.079 | 0.081 | 0.079 |  

0.088 | 0.170 | 0.086 | 0.079 | 0.327 |  

##          | 0.378 | 0.381 | 0.376 | 0.344 | 0.328 | 0.265 |  

0.248 | 0.360 | 0.292 | 0.298 |  

##          | 0.040 | 0.039 | 0.031 | 0.026 | 0.027 | 0.026 |  

0.029 | 0.056 | 0.028 | 0.026 |  

## -----|-----|-----|-----|-----|-----|-----|  

##          Modéré |   23 |   34 |   15 |   20 |   19 |   14 |  

23 |   38 |   16 |   14 |   216 |  

##          | 0.001 | 6.345 | 0.492 | 0.898 | 0.115 | 2.308 |  

0.196 | 0.643 | 1.090 | 1.168 |  

##          | 0.106 | 0.157 | 0.069 | 0.093 | 0.088 | 0.065 |  

0.106 | 0.176 | 0.074 | 0.065 | 0.099 |  

##          | 0.099 | 0.152 | 0.083 | 0.123 | 0.107 | 0.066 |  

0.091 | 0.113 | 0.077 | 0.074 |  

##          | 0.011 | 0.016 | 0.007 | 0.009 | 0.009 | 0.006 |  

0.011 | 0.017 | 0.007 | 0.006 |  

## -----|-----|-----|-----|-----|-----|-----|  

##          Non anémique |   122 |   104 |   98 |   87 |   100 |   141 |  

168 |   177 |   132 |   118 |   1247 |  

##          | 1.005 | 4.450 | 0.321 | 0.446 | 0.022 | 3.315 |  

3.437 | 1.270 | 1.237 | 0.968 |  

##          | 0.098 | 0.083 | 0.079 | 0.070 | 0.080 | 0.113 |  

0.135 | 0.142 | 0.106 | 0.095 | 0.573 |  

##          | 0.524 | 0.466 | 0.541 | 0.534 | 0.565 | 0.668 |  

0.661 | 0.527 | 0.632 | 0.628 |  

##          | 0.056 | 0.048 | 0.045 | 0.040 | 0.046 | 0.065 |  

0.077 | 0.081 | 0.061 | 0.054 |  

## -----|-----|-----|-----|-----|-----|-----|

```

```

-----|-----|-----|-----|-----|
##   Column Total |      233 |      223 |      181 |      163 |      177 |      211 |
254 |      336 |      209 |      188 |     2175 |          |          |
##           |      0.107 |      0.103 |      0.083 |      0.075 |      0.081 |      0.097 |
0.117 |      0.154 |      0.096 |      0.086 |          |          |          |
## -----|-----|-----|-----|-----|-----|-----|
-----|-----|-----|-----|-----|
##  

##  

## Statistics for All Table Factors  

##  

##  

## Pearson's Chi-squared test  

## -----  

## Chi^2 = 44.68585      d.f. = 18      p = 0.0004599718  

##  

##  

##  

##
```

En examinant le graphique ci-dessus, nous pouvons constater que le pourcentage d'enfants anémiques varie d'une région à l'autre. De même, le pourcentage des cas modérés présente des variations régionales. La région qui affiche le taux le plus élevé d'enfants anémiques est la région Central, avec un pourcentage de 53%, soit plus de la moitié. En revanche, les régions Ashanti et Brong-Ahafo enregistrent les taux les plus bas d'enfants anémiques (33% et 34%) ainsi que des pourcentages équivalents de cas modérés. Ces deux régions centrales présentent des pourcentages quasiment identiques et sont situées à proximité l'une de l'autre.

Par ailleurs, on observe que les régions Upper West et Upper East présentent des taux similaires d'enfants anémiques, ainsi que des pourcentages identiques de cas modérés. Les régions Western, Central, Greater Accra, Volta et Eastern, situées au sud du Ghana, présentent également des pourcentages presque similaires de cas anémiques.

Selon les analyses précédentes, il semble y avoir une corrélation entre l'anémie chez les enfants et la région du ménage. Afin de vérifier cette relation, nous avons effectué un test du Khi-deux d'indépendance en formulant deux hypothèses :

Hypothèse nulle : les deux variables sont indépendantes. Hypothèse alternative : les deux variables ne sont pas indépendantes.

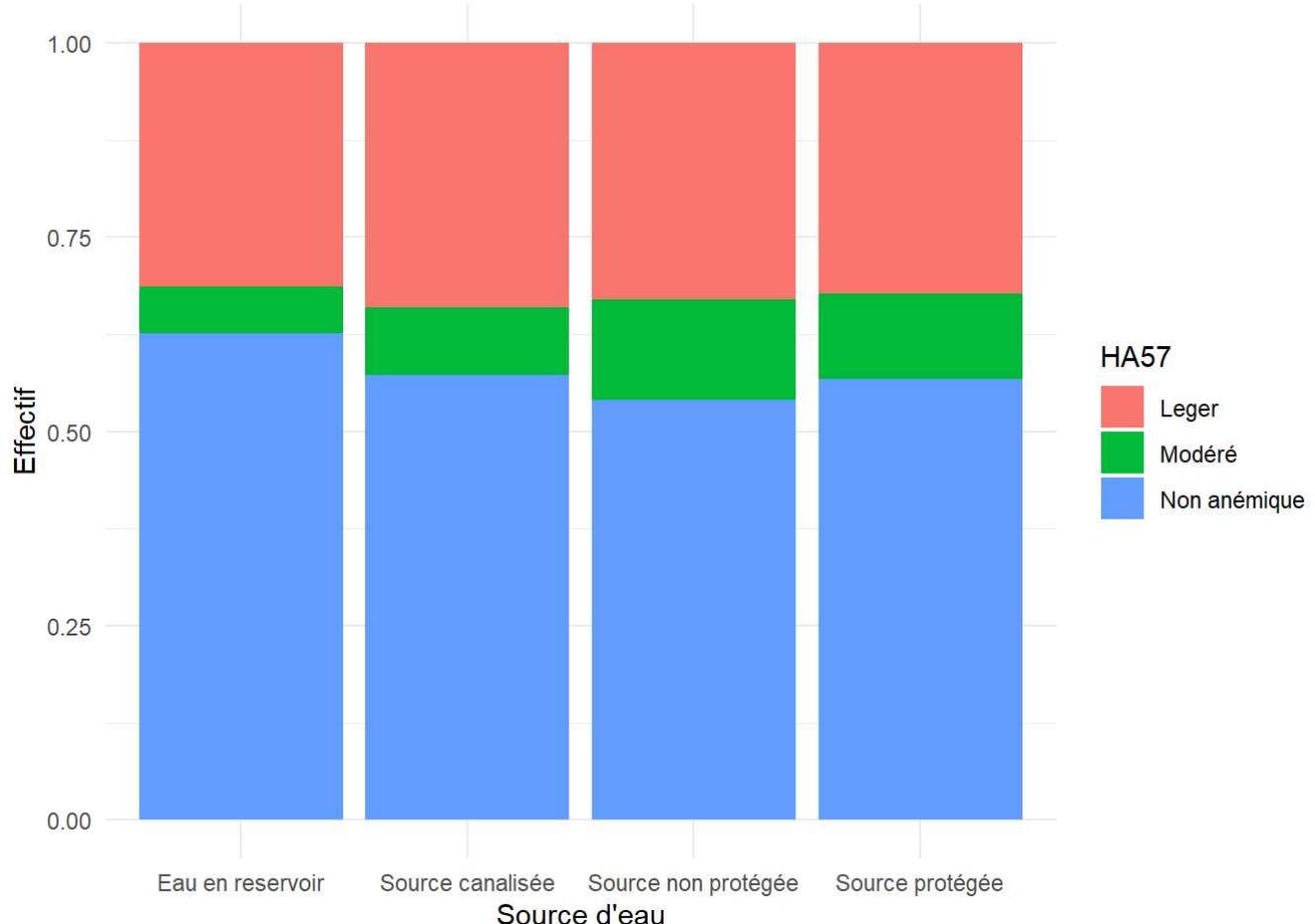
Les résultats du test du Khi-2 révèlent une p-valeur de 0,5 %, ce qui est pratiquement négligeable. Par conséquent, nous rejetons l'hypothèse d'indépendance entre les deux variables.

Croisement des variables anémie et source d'eau

La qualité de l'eau et les composantes chimiques varient d'une source à l'autre. Pour mener une analyse approfondie de cette variable et effectuer des regroupements pertinents, il serait nécessaire de disposer des caractéristiques spécifiques de chaque source. Malheureusement, ces informations ne nous sont pas accessibles en raison d'un manque d'informations détaillées. Néanmoins, nous procédons à des regroupements en fonction de critères basiques tels que la source protégée, non protégée, la source canalisée et l'eau en réservoir.

```
anemie_source <- table(enf$cHA57,enf$cHV201)

ggplot(enf) +
  aes(x = cHV201, fill = HA57) +
  geom_bar(position="fill") +
  scale_fill_hue(direction = 1) +
  labs(x = "Source d'eau",y="Effectif")+
  theme_minimal()
```



```
tab_source <- data.frame(ID=c(11,12,13,21,31,32,41,42,43,51,61,71,72), region=c(
  "Canalisée dans le logement",
  "Canalisée dans la cour/parcelle",
  "Robinet public/fontaine",
  "Puits tubulaire ou forage",
  "Puits protégé",
  "Puits non protégé",
  "Source protégée",
  "Source non protégée",
  "Rivière/barrage/lac/étang/cours d'eau",
  "Eau de pluie",
  "Camion-citerne",
  "Eau en bouteille",
  "Eau en sachet") )

tab_source
```

```

##   ID          region
## 1 11      Canalisée dans le logement
## 2 12      Canalisée dans la cour/parcelle
## 3 13      Robinet public/fontaine
## 4 21      Puits tubulaire ou forage
## 5 31      Puits protégé
## 6 32      Puits non protégé
## 7 41      Source protégée
## 8 42      Source non protégée
## 9 43 Rivièvre/barrage/lac/étang/cours d'eau
## 10 51     Eau de pluie
## 11 61     Camion-citerne
## 12 71     Eau en bouteille
## 13 72     Eau en sachet

```

var_source

```

## # A tibble: 4 × 5
##   cHV201      effectifs eff_cum pourcent pour_cum
##   <fct>       <int>    <int>     <dbl>     <dbl>
## 1 Eau en réservoir      310      310     14.2     14.2
## 2 Source canalisée      608      918     28.0     42.2
## 3 Source non protégée  327     1245     15.0     57.2
## 4 Source protégée       930     2175     42.8    100.

```

```
CrossTable(enf$anemie,enf$cHV201,chisq = T, dnn = c("Niveau d'anémie","source d'eau"))
```

```

##  

##  

##      Cell Contents  

## |-----|  

## |           N |  

## | Chi-square contribution |  

## |           N / Row Total |  

## |           N / Col Total |  

## |           N / Table Total |  

## |-----|  

##  

##  

## Total Observations in Table: 2175  

##  

##  

##          | source d'eau  

## Niveau d'anémie |   Eau en réservoir |   Source canalisée |   Source non protégée | So  

urce protégée |           Row Total |  

## -----|-----|-----|-----|-----|  

##  

##     Anémique |           116 |           260 |           150 |  

402 |           928 |  

##           |           2.001 |           0.001 |           0.787 |  

0.068 |  

##           |           0.125 |           0.280 |           0.162 |  

0.433 |           0.427 |  

##           |           0.374 |           0.428 |           0.459 |  

0.432 |  

##           |           0.053 |           0.120 |           0.069 |  

0.185 |  

## -----|-----|-----|-----|-----|  

##  

##     Non anémique |           194 |           348 |           177 |  

528 |           1247 |  

##           |           1.489 |           0.001 |           0.586 |  

0.051 |  

##           |           0.156 |           0.279 |           0.142 |  

0.423 |           0.573 |  

##           |           0.626 |           0.572 |           0.541 |  

0.568 |  

##           |           0.089 |           0.160 |           0.081 |  

0.243 |  

## -----|-----|-----|-----|-----|  

##  

##     Column Total |           310 |           608 |           327 |  

930 |           2175 |  

##           |           0.143 |           0.280 |           0.150 |  

0.428 |  

## -----|-----|-----|-----|-----|  

##  

##  

## Statistics for All Table Factors  

##  

##  

##
```

```

## Pearson's Chi-squared test
## -----
## Chi^2 = 4.983508      d.f. = 3      p = 0.1730087
## 
## 
## 
```

Les graphiques ci-dessus révèlent que le taux d'enfants anémiques est sensiblement similaire quel que soit la source d'eau. Cependant, on observe une aggravation de l'anémie chez les enfants consommant de l'eau provenant d'une source non protégée, tandis qu'une diminution est constatée chez ceux buvant de l'eau issue d'un réservoir (bouteille, sachet, camion-citerne).

Suite à l'application du test du Khi-2, nous obtenons une p-valeur de 17%. Cette valeur, bien qu'elle ne renforce pas l'indépendance entre les deux variables, ne permet pas non plus de rejeter complètement l'hypothèse de dépendance.

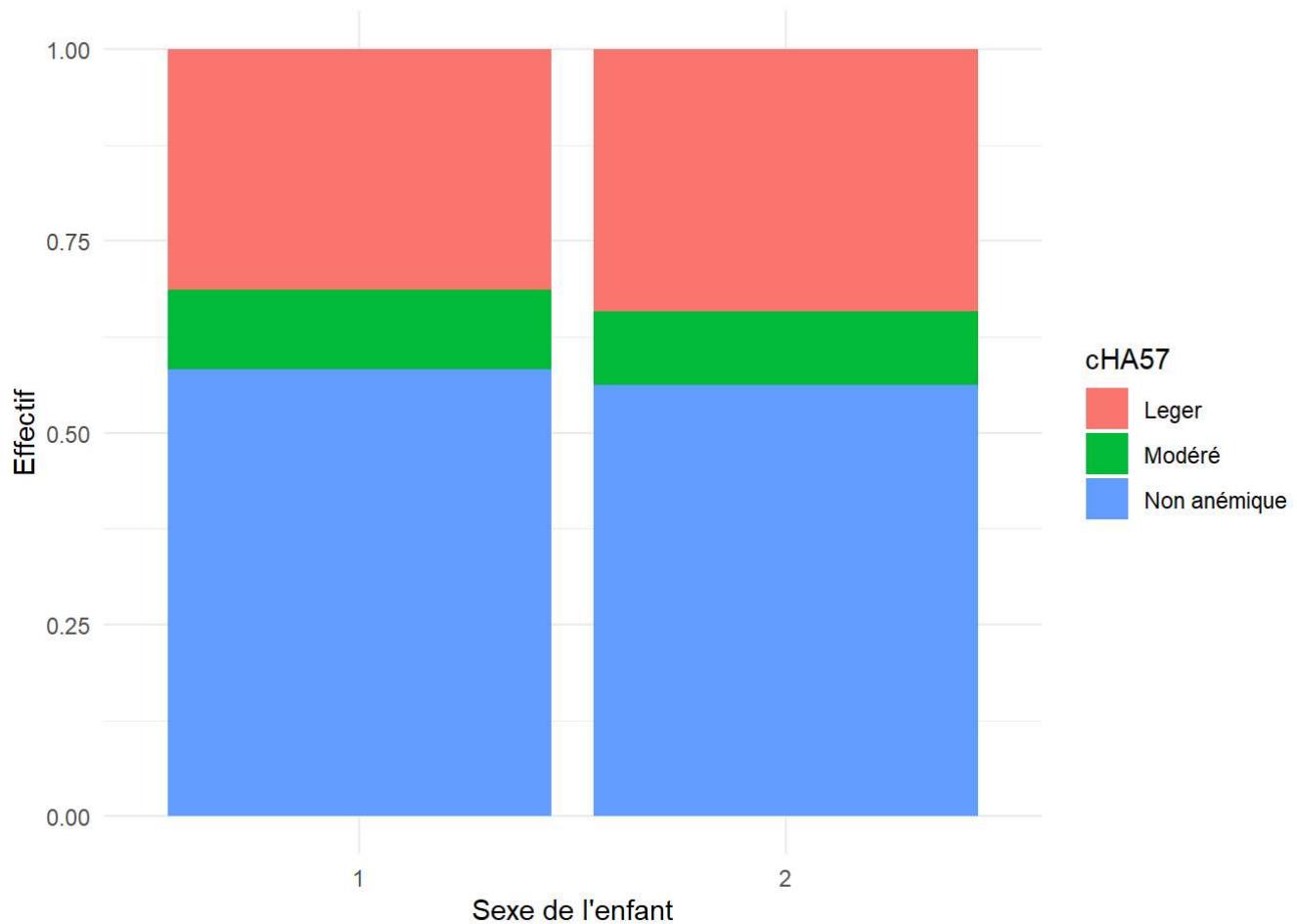
Croisement des variables anémie et sexe de l'enfant

La sélection de cette variable a été motivée par la curiosité. Nous souhaitons déterminer si la maladie est spécifique à un seul genre. Pour ce faire, nous avons effectué les analyses ci-dessous.

```

anemie_sex <- table(enf$cHA57,enf$cHC27)

ggplot(enf) +
  aes(x = cHC27, fill = cHA57) +
  geom_bar(position="fill") +
  scale_fill_hue(direction = 1) +
  labs(x = "Sexe de l'enfant",y="Effectif")+
  theme_minimal() 
```



```
CrossTable(enf$cHA57,enf$cHC27,chisq = T, dnn = c("Niveau d'anémie","Sexe de l'enfant"))
```

```

## 
## 
##   Cell Contents
## |-----|
## |           N |
## | Chi-square contribution |
## |           N / Row Total |
## |           N / Col Total |
## |           N / Table Total |
## |-----|
## 
## 
## Total Observations in Table:  2175
## 
## 
##          | Sexe de l'enfant
## Niveau d'anémie |      1 |      2 | Row Total |
## -----|-----|-----|-----|
##     Leger | 352 | 360 | 712 |
##             | 0.610 | 0.649 | 
##             | 0.494 | 0.506 | 0.327 |
##             | 0.314 | 0.342 | 
##             | 0.162 | 0.166 | 
## -----|-----|-----|-----|
##     Modéré | 115 | 101 | 216 |
##             | 0.121 | 0.129 | 
##             | 0.532 | 0.468 | 0.099 |
##             | 0.103 | 0.096 | 
##             | 0.053 | 0.046 | 
## -----|-----|-----|-----|
##     Non anémique | 654 | 593 | 1247 |
##             | 0.198 | 0.211 | 
##             | 0.524 | 0.476 | 0.573 |
##             | 0.583 | 0.563 | 
##             | 0.301 | 0.273 | 
## -----|-----|-----|-----|
##     Column Total | 1121 | 1054 | 2175 |
##             | 0.515 | 0.485 | 
## -----|-----|-----|-----|
## 
## 
## Statistics for All Table Factors
## 
## 
## Pearson's Chi-squared test
## -----|
## Chi^2 = 1.91917    d.f. = 2    p = 0.3830519
## 
## 
## 

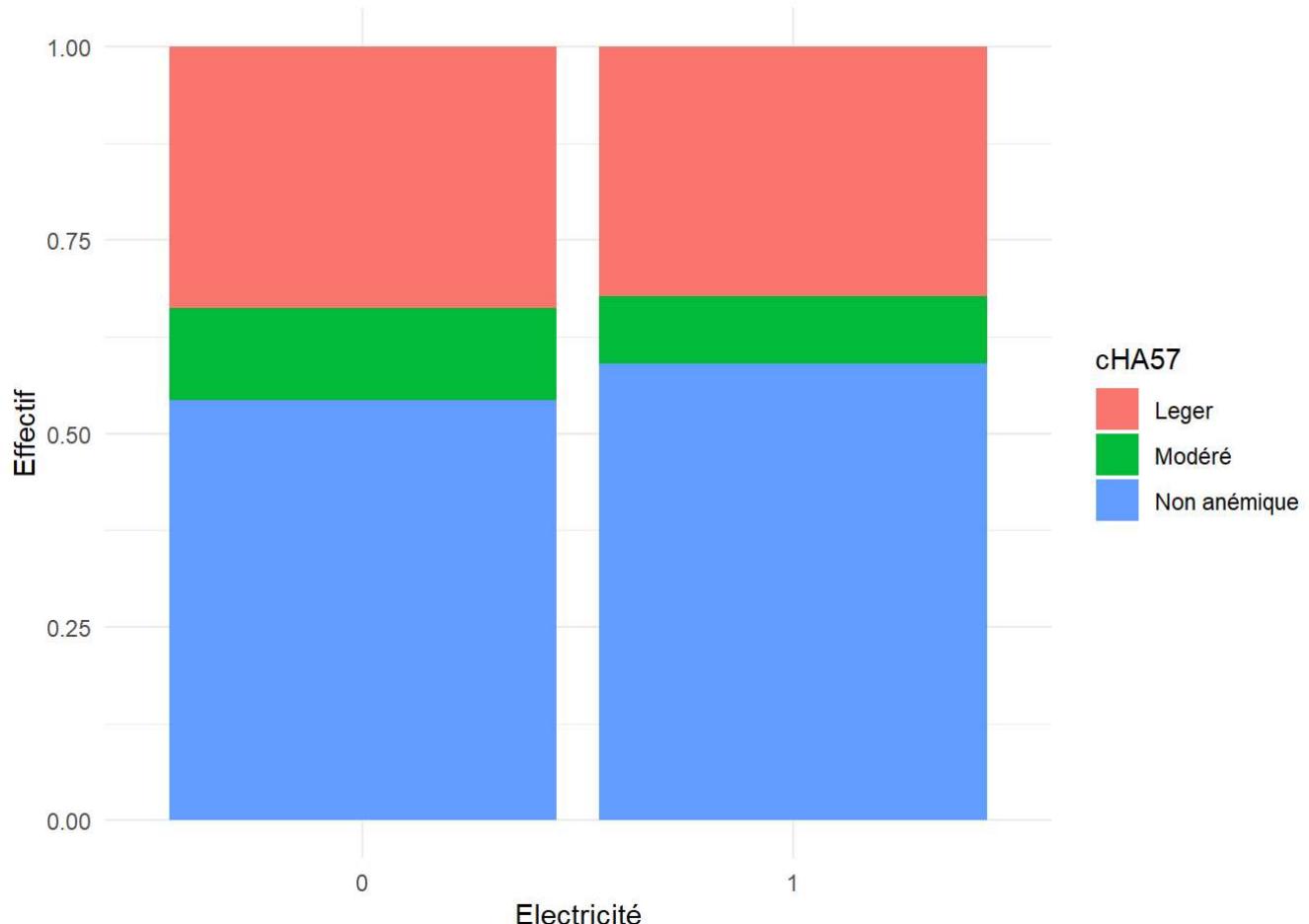
```

Nous constatons que la répartition des niveaux d'anémie chez les enfants est pratiquement similaire pour les deux sexes. En utilisant les résultats du test du khi-2, on peut conclure que l'anémie ne dépend pas du genre de l'enfant, ce qui est en accord avec les résultats de l'analyse univariée.

Croisement des variables anémie et électricité

```
anemie_elec <- table(enf$cHA57,enf$cHV206)
```

```
ggplot(enf) +
  aes(x = cHV206, fill = cHA57) +
  geom_bar(position="fill") +
  scale_fill_hue(direction = 1) +
  labs(x = "Électricité",y="Effectif")+
  theme_minimal()
```



```
CrossTable(enf$cHA57,enf$cHV206,chisq = T, dnn = c("Niveau d'anémie","Électricité"))
```

```

## 
## 
##   Cell Contents
## |-----|
## |           N |
## | Chi-square contribution |
## |           N / Row Total |
## |           N / Col Total |
## |           N / Table Total |
## |-----|
## 
## 
## Total Observations in Table:  2175
## 
## 
##          | Electricité
## Niveau d'anémie |     0 |      1 | Row Total |
## -----|-----|-----|-----|
##       Leger | 268 | 444 | 712 |
##             | 0.231 | 0.133 | 
##             | 0.376 | 0.624 | 0.327 |
##             | 0.337 | 0.322 | 
##             | 0.123 | 0.204 | 
## -----|-----|-----|-----|
##       Modéré | 95 | 121 | 216 |
##             | 3.262 | 1.879 | 
##             | 0.440 | 0.560 | 0.099 |
##             | 0.119 | 0.088 | 
##             | 0.044 | 0.056 | 
## -----|-----|-----|-----|
## Non anémique | 432 | 815 | 1247 |
##             | 1.243 | 0.716 | 
##             | 0.346 | 0.654 | 0.573 |
##             | 0.543 | 0.591 | 
##             | 0.199 | 0.375 | 
## -----|-----|-----|-----|
## Column Total | 795 | 1380 | 2175 |
##             | 0.366 | 0.634 | 
## -----|-----|-----|-----|
## 
## 
## Statistics for All Table Factors
## 
## 
## Pearson's Chi-squared test
## -----|
## Chi^2 = 7.463897    d.f. = 2    p = 0.02394613
## 
## 
## 

```

Nous constatons que dans les foyers dépourvus d'électricité, le taux d'enfants anémiques est plus élevé que celui des foyers disposant de l'électricité. Il est à noter également que le taux de cas modérés est plus élevé dans les ménages sans électricité. La relation entre ces deux variables est confirmée par le test du Khi-2. En

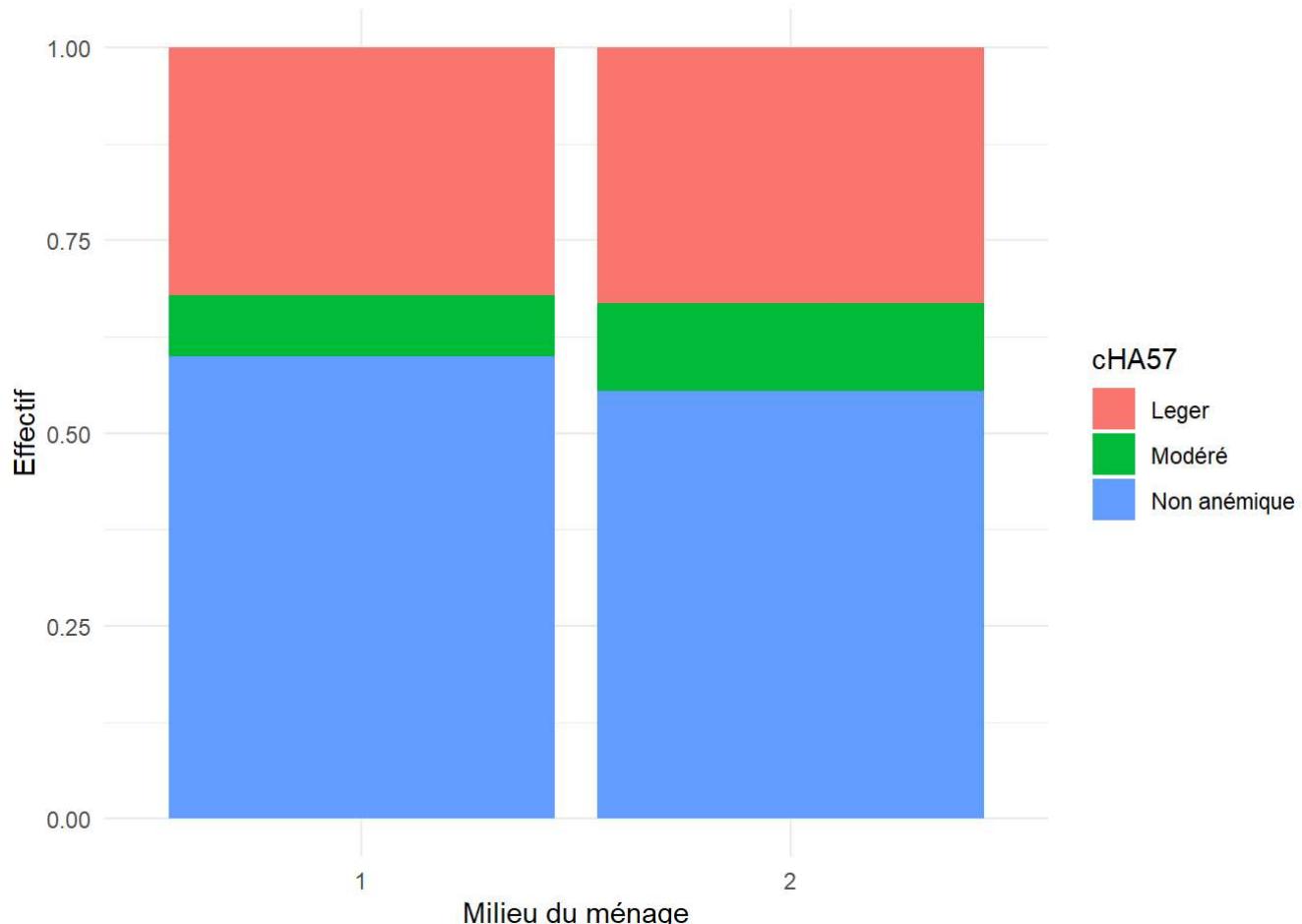
effet, la p-valeur (2,4%) obtenue à partir des résultats du test nous permet de rejeter l'hypothèse d'indépendance entre ces deux variables.

Croisement des variables anémie et milieu du ménage

Il est également intéressant de croiser ces deux variables, car les zones rurales et urbaines présentent des caractéristiques distinctes, telles que les conditions environnementales et les habitudes alimentaires. Ces facteurs pourraient contribuer à des disparités de taux d'enfants anémiques entre les deux zones.

```
anemie_zone <- table(enf$cHA57,enf$cHV025)

ggplot(enf) +
  aes(x = cHV025, fill = cHA57) +
  geom_bar(position="fill") +
  scale_fill_hue(direction = 1) +
  labs(x = "Milieu du ménage",y="Effectif")+
  theme_minimal()
```



```
CrossTable(enf$cHA57,enf$cHV025,chisq = T, dnn = c("Niveau d'anémie","Milieu du ménage"))
```

```

## 
## 
##   Cell Contents
## |-----|
## |           N |
## | Chi-square contribution |
## |           N / Row Total |
## |           N / Col Total |
## |           N / Table Total |
## |-----|
## 
## 
## Total Observations in Table:  2175
## 
## 
##          | Milieu du ménage
## Niveau d'anémie |      1 |      2 | Row Total |
## -----|-----|-----|-----|
##     Leger |    286 |    426 |    712 |
##             | 0.087 | 0.060 |      |
##             | 0.402 | 0.598 | 0.327 |
##             | 0.322 | 0.331 |      |
##             | 0.131 | 0.196 |      |
## -----|-----|-----|-----|
##     Modéré |     70 |    146 |    216 |
##             | 3.788 | 2.618 |      |
##             | 0.324 | 0.676 | 0.099 |
##             | 0.079 | 0.114 |      |
##             | 0.032 | 0.067 |      |
## -----|-----|-----|-----|
## Non anémique |    533 |    714 |   1247 |
##             | 1.066 | 0.737 |      |
##             | 0.427 | 0.573 | 0.573 |
##             | 0.600 | 0.555 |      |
##             | 0.245 | 0.328 |      |
## -----|-----|-----|-----|
## Column Total |    889 |   1286 |   2175 |
##             | 0.409 | 0.591 |      |
## -----|-----|-----|-----|
## 
## 
## Statistics for All Table Factors
## 
## 
## Pearson's Chi-squared test
## -----|
## Chi^2 = 8.355141      d.f. = 2      p = 0.01533572
## 
## 
## 
## 
```

D'après les graphiques précédents, on observe une différence de taux d'enfants anémiques entre les deux zones (rurales et urbaines). Notamment, le pourcentage de cas modérés est plus élevé en milieu rural qu'en milieu urbain.

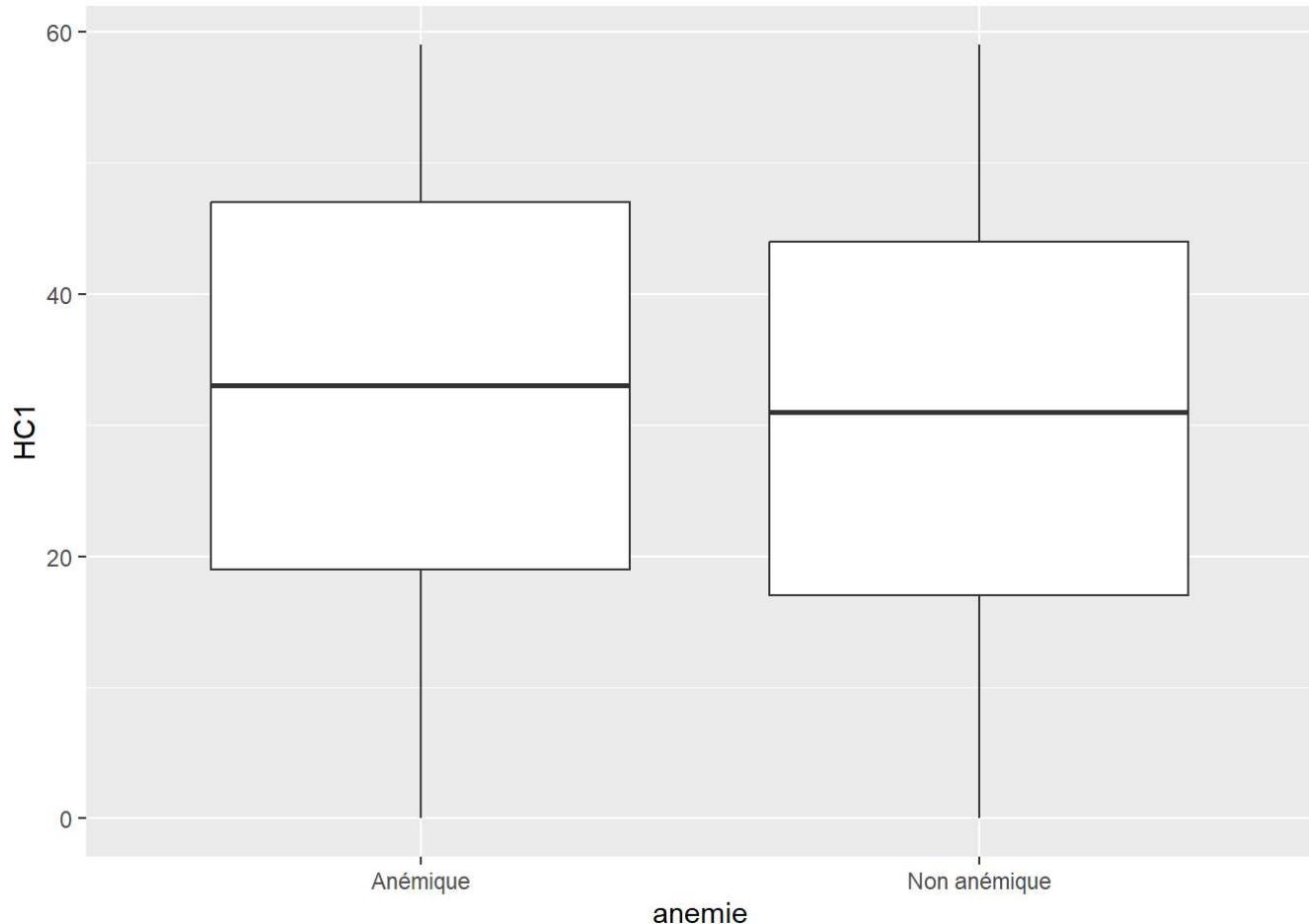
Il semble y avoir une corrélation entre la zone et le niveau d'anémie des enfants, ce qui est corroboré par le test du khi-2 donnant une p-valeur de 1,5%.

Croisement des variables anémie et âge

Il s'agit ici d'analyser si l'anémie affecte davantage certaines tranches d'âge que d'autres. Selon les diagrammes en boîte ci-dessous, on constate que la distribution de l'âge pour les enfants anémiques est presque identique à celle des enfants non anémiques. Ainsi, on ne peut pas conclure à l'existence d'une dépendance entre ces deux variables.

```
anemie_age <- table(enf$anemie,enf$HC1)

ggplot(enf,aes(x =anemie, y= HC1)) +
  geom_boxplot()
```



ACM:

Dans cette section, nous entreprenons une analyse des correspondances multiples sur les variables d'intérêt, auxquelles nous avons ajouté les variables nécessaires. Dans un premier temps, nous examinons la table récapitulative des variables ajoutées aux variables de départ pour traiter les cas de valeurs manquantes. Nous avons décidé de conserver 15 variables, en éliminant 6 parmi celles demandées que nous jugeons moins importantes pour notre étude.

En particulier, les variables radio (cHV207) et télévision (cHV208) remplissent des rôles presque similaires (transmission d'informations, sensibilisation), et la plupart des ménages utilisent davantage la télévision que la radio, même s'ils possèdent les deux. C'est pourquoi nous avons choisi de conserver la variable télévision plutôt que la radio. Quant aux variables matière de construction du sol (cHV213), présence d'un PC ou d'une

tablette (cSH110L) et présence d'une armoire ou d'un placard (cSH110R), nous ne les avons pas conservées car nous estimons qu'elles ne sont pas directement liées à la maladie d'anémie chez les enfants. En ce qui concerne les variables moustiquaires autour des lits (cHV227), l'âge des enfants (HC1) et le sexe des enfants (cHC27), les résultats obtenus à partir des analyses bivariées nous ont conduit à rejeter ces variables. Cela justifie nos choix quant aux variables que nous utiliserons ultérieurement.

Dans la table "menage1", nous stockons les variables que nous avons choisi de conserver.

```
menage1 <- select(enf, anemie, cHC61, cHV024, cHV025, cHV227, cHV270, cHV206, cHV201, cHV205, cHV208, cHV209, cHV226, cSH110P, cSH110Q, cSH110S, HC1)

summary(select(enf, cHV205, cHV208, cHV209, cHV226, cSH110P, cSH110Q, cSH110S))
```

```
##      cHV205      cHV208      cHV209      cHV226      cSH110P      cSH110Q 
## 31     :692    0     :1072   0:1606     8     :1245    0: 561    0     : 476 
## 21     :669    1     :1102   1: 569     7     : 593    1:1614    1     :1698 
## 22     :358  NA's:    1                  2     : 308                  NA's:    1 
## 23     :177                  9     : 16 
## 12     :163                  1     : 5 
## 11     : 65                  6     : 3 
## (Other): 51                (Other): 5 
## cSH110S 
## 0:1985 
## 1: 190 
## 
## 
## 
## 
##
```

Traitement des valeurs manquantes:

Nous observons que les valeurs manquantes sont négligeables, avec seulement une occurrence manquante pour les variables cHV208 et cSH110Q. Nous avons choisi de remplacer ces cas par la modalité la plus fréquente pour chacune des variables.

```
menage1$cHV208[is.na(menage1$cHV208)] <- 1

menage1$cSH110Q[is.na(menage1$cSH110Q)] <- 1

menage1$cHV208 <- as.factor(menage1$cHV208)

menage1$cSH110Q <- as.factor(menage1$cSH110Q)

summary(select(menage1, cHV205, cHV208, cHV209, cHV226, cSH110P, cSH110Q, cSH110S))
```

```

##      cHV205    cHV208    cHV209      cHV226    cSH110P    cSH110Q    cSH110S
## 31     :692    0:1072   0:1606      8       :1245    0: 561    0: 476    0:1985
## 21     :669    1:1103   1: 569      7       : 593    1:1614   1:1699    1: 190
## 22     :358          2       : 308
## 23     :177          9       : 16
## 12     :163          1       : 5
## 11     : 65          6       : 3
## (Other): 51          (Other): 5

```

Certaines variables, telles que la région des enfants de moins de 5 ans, le type de toilettes et le type de combustible de cuisson pour les ménages, nécessitent un regroupement et un recodage des modalités afin de faciliter les calculs de l'Analyse des Correspondances Multiples (ACM) et sa compréhension.

Regroupement des modalités

```
class(menage1$cHV205)
```

```
## [1] "factor"
```

```
menage1$cHV205<-fct_collapse(menage1$cHV205,"aucune installation"= c("31"), "latrine" = c("21","22","23"),"chasse d'eau"=c("12","11"),"autre toilette"=c("13","14","15","42","43"))
```

```
summary(menage1$cHV226)
```

	1	2	5	6	7	8	9	10	95	
##	5	308		1	3	593	1245	16	1	3

```
menage1$cHV226<-fct_collapse(menage1$cHV226,"charbon"=c("6","7"),"bois"=c("8","9"),"elec_gaz"=c("1","2","5","10","95"))
summary(menage1$cHV226)
```

	elec_gaz	charbon	bois
##	318	596	1261

```
menage1$cHV024<-fct_collapse(menage1$cHV024,"Nord"=c("10","9"),"centre"=c("6","7"), "sud-ouest"=c("1","2"),"sud-est"=c("3","4","5"),"Northem"="8")
summary(menage1$cHV024)
```

	sud-ouest	sud-est	centre	Northem	Nord
##	456	521	465	336	397

```
menage1$HC1 <- as.numeric(menage1$HC1)
menage1$HC1<-cut(menage1$HC1, breaks = 4,labels=c("0-15","15-30","30-45","45-60"))
```

Récodage des modalités

```

menage1$cHC61 <- fct_recode(menage1$cHC61, "pas etud" = "0", "prim" = "1", "colg" = "2", "Lyc" = "3")
menage1$cHV025 <- fct_recode(menage1$cHV025, "urbain" = "1", "rural" = "2")
menage1$cHV270 <- fct_recode(menage1$cHV270, "t_pauvre" = "1", "pauvre" = "2", "modeste" = "3", "rich e" = "4", "t_riche" = "5")
menage1$cHV208 <- fct_recode(menage1$cHV208, "non" = "0", "oui" = "1")
#menage1$cHC27 <- fct_recode(menage1$cHC27, "M" = "1", "F" = "2")

menage1$cHV209 <- fct_recode(menage1$cHV209, "non" = "0", "oui" = "1")
menage1$cHV206 <- fct_recode(menage1$cHV206, "non" = "0", "oui" = "1")
menage1$cHV227 <- fct_recode(menage1$cHV227, "non" = "0", "oui" = "1")

menage1$cSH110P <- fct_recode(menage1$cSH110P, "non" = "0", "oui" = "1")
menage1$cSH110Q <- fct_recode(menage1$cSH110Q, "non" = "0", "oui" = "1")
menage1$cSH110S <- fct_recode(menage1$cSH110S, "non" = "0", "oui" = "1")

# renommons les colonnes
names(menage1) <- c("Anemie", "Educ_M", "Reg", "Zn", "Mous", "Irish", "Elec", "EP", "Toil", "Tlv", "Ref", "Comb", "Lit", "Tab", "Net", "age_enf")

menage2 <- select(menage1, Zn, EP, Toil, Tlv, Ref, Lit, Tab, Comb, Net)

# Visualisation de la table
view(menage2)

```

ACM du tableau de Burt

```

library(ade4)
menage2 <- data.frame(menage2)
acm.burt(menage2, menage2)

```

	Zn.urbain	Zn.rural	EP.Eau en reservoir
## Zn.urbain	889	0	251
## Zn.rural	0	1286	59
## EP.Eau en reservoir	251	59	310
## EP.Source canalisée	403	205	0
## EP.Source non protégée	45	282	0
## EP.Source protégée	190	740	0
## Toil.chasse d'eau	207	21	101
## Toil.autre toilette	44	7	22
## Toil.latrine	518	686	171
## Toil.aucune installation	120	572	16
## Tlv.non	191	881	32
## Tlv.oui	698	405	278
## Ref.non	459	1147	108
## Ref.oui	430	139	202
## Lit.non	115	446	25
## Lit.oui	774	840	285
## Tab.non	117	359	29
## Tab.oui	772	927	281
## Comb.elec_gaz	265	53	162
## Comb.charbon	433	163	135
## Comb.bois	191	1070	13
## Net.non	736	1249	239
## Net.oui	153	37	71
## EP.Source canalisée	EP.Source non protégée		
## Zn.urbain	403	45	
## Zn.rural	205	282	
## EP.Eau en reservoir	0	0	
## EP.Source canalisée	608	0	
## EP.Source non protégée	0	327	
## EP.Source protégée	0	0	
## Toil.chasse d'eau	106	2	
## Toil.autre toilette	18	3	
## Toil.latrine	394	166	
## Toil.aucune installation	90	156	
## Tlv.non	179	263	
## Tlv.oui	429	64	
## Ref.non	372	305	
## Ref.oui	236	22	
## Lit.non	97	126	
## Lit.oui	511	201	
## Tab.non	70	105	
## Tab.oui	538	222	
## Comb.elec_gaz	116	4	
## Comb.charbon	276	19	
## Comb.bois	216	304	
## Net.non	538	323	
## Net.oui	70	4	
## EP.Source protégée	Toil.chasse d'eau		
## Zn.urbain	190	207	
## Zn.rural	740	21	
## EP.Eau en reservoir	0	101	
## EP.Source canalisée	0	106	
## EP.Source non protégée	0	2	
## EP.Source protégée	930	19	

## Toil.chasse d'eau	19	228		
## Toil.autre toilette	8	0		
## Toil.latrine	473	0		
## Toil.aucune installation	430	0		
## Tlv.non	598	15		
## Tlvoui	332	213		
## Ref.non	821	62		
## Ref.oui	109	166		
## Lit.non	313	12		
## Lit.oui	617	216		
## Tab.non	272	20		
## Tab.oui	658	208		
## Comb.elec_gaz	36	130		
## Comb.charbon	166	89		
## Comb.bois	728	9		
## Net.non	885	155		
## Net.oui	45	73		
##	Toil.autre toilette	Toil.latrine		
## Zn.urbain	44	518		
## Zn.rural	7	686		
## EP.Eau en reservoir	22	171		
## EP.Source canalisée	18	394		
## EP.Source non protégée	3	166		
## EP.Source protégée	8	473		
## Toil.chasse d'eau	0	0		
## Toil.autre toilette	51	0		
## Toil.latrine	0	1204		
## Toil.aucune installation	0	0		
## Tlv.non	8	531		
## Tlv.oui	43	673		
## Ref.non	23	877		
## Ref.oui	28	327		
## Lit.non	4	220		
## Lit.oui	47	984		
## Tab.non	6	183		
## Tab.oui	45	1021		
## Comb.elec_gaz	17	154		
## Comb.charbon	26	396		
## Comb.bois	8	654		
## Net.non	44	1108		
## Net.oui	7	96		
##	Toil.aucune installation	Tlv.non	Tlv.oui	Ref.non
## Zn.urbain	120	191	698	459
## Zn.rural	572	881	405	1147
## EP.Eau en reservoir	16	32	278	108
## EP.Source canalisée	90	179	429	372
## EP.Source non protégée	156	263	64	305
## EP.Source protégée	430	598	332	821
## Toil.chasse d'eau	0	15	213	62
## Toil.autre toilette	0	8	43	23
## Toil.latrine	0	531	673	877
## Toil.aucune installation	692	518	174	644
## Tlv.non	518	1072	0	1041
## Tlv.oui	174	0	1103	565
## Ref.non	644	1041	565	1606
## Ref.oui	48	31	538	0

## Lit.non		325	458	103	538
## Lit.oui		367	614	1000	1068
## Tab.non		267	349	127	438
## Tab.oui		425	723	976	1168
## Comb.elec_gaz		17	19	299	76
## Comb.charbon		85	139	457	358
## Comb.bois		590	914	347	1172
## Net.non		678	1056	929	1554
## Net.oui		14	16	174	52
##	Ref.oui	Lit.non	Lit.oui	Tab.non	Tab.oui
## Zn.urbain	430	115	774	117	772
## Zn.rural	139	446	840	359	927
## EP.Eau en reservoir	202	25	285	29	281
## EP.Source canalisée	236	97	511	70	538
## EP.Source non protégée	22	126	201	105	222
## EP.Source protégée	109	313	617	272	658
## Toil.chasse d'eau	166	12	216	20	208
## Toil.autre toilette	28	4	47	6	45
## Toil.latrine	327	220	984	183	1021
## Toil.aucune installation	48	325	367	267	425
## Tlv.non	31	458	614	349	723
## Tlv.oui	538	103	1000	127	976
## Ref.non	0	538	1068	438	1168
## Ref.oui	569	23	546	38	531
## Lit.non	23	561	0	317	244
## Lit.oui	546	0	1614	159	1455
## Tab.non	38	317	159	476	0
## Tab.oui	531	244	1455	0	1699
## Comb.elec_gaz	242	17	301	18	300
## Comb.charbon	238	93	503	88	508
## Comb.bois	89	451	810	370	891
## Net.non	431	551	1434	464	1521
## Net.oui	138	10	180	12	178
##	Comb.charbon	Comb.bois	Net.non	Net.oui	
## Zn.urbain	433	191	736	153	
## Zn.rural	163	1070	1249	37	
## EP.Eau en reservoir	135	13	239	71	
## EP.Source canalisée	276	216	538	70	
## EP.Source non protégée	19	304	323	4	
## EP.Source protégée	166	728	885	45	
## Toil.chasse d'eau	89	9	155	73	
## Toil.autre toilette	26	8	44	7	
## Toil.latrine	396	654	1108	96	
## Toil.aucune installation	85	590	678	14	
## Tlv.non	139	914	1056	16	
## Tlv.oui	457	347	929	174	
## Ref.non	358	1172	1554	52	
## Ref.oui	238	89	431	138	
## Lit.non	93	451	551	10	
## Lit.oui	503	810	1434	180	
## Tab.non	88	370	464	12	
## Tab.oui	508	891	1521	178	
## Comb.elec_gaz	0	0	222	96	
## Comb.charbon	596	0	538	58	
## Comb.bois	0	1261	1225	36	

	538	1225	1985	0
## Net.non	58	36	0	190

```
library(FactoMineR)
```

```
##
## Attachement du package : 'FactoMineR'
```

```
## L'objet suivant est masqué depuis 'package:ade4':
##
##     reconst
```

```
acm1<-MCA(menage2,graph=FALSE, method = "Burt")
acm1$eig
```

	eigenvalue	percentage of variance	cumulative percentage of variance
## dim 1	0.199871072	64.1631484	64.16315
## dim 2	0.026310111	8.4461424	72.60929
## dim 3	0.017947200	5.7614585	78.37075
## dim 4	0.013149413	4.2212599	82.59201
## dim 5	0.012693282	4.0748315	86.66684
## dim 6	0.009505299	3.0514167	89.71826
## dim 7	0.007811017	2.5075138	92.22577
## dim 8	0.006732469	2.1612753	94.38705
## dim 9	0.004917665	1.5786819	95.96573
## dim 10	0.003656827	1.1739243	97.13965
## dim 11	0.003044097	0.9772241	98.11688
## dim 12	0.002419275	0.7766423	98.89352
## dim 13	0.002232367	0.7166405	99.61016
## dim 14	0.001214371	0.3898406	100.00000

Choix des axes:

Valeurs propres:

Nous avons pour cette ACM s=9 variables (questions) actives avec, en tout, p=23 modalités. Pour cette analyse, le nombre d'axes ainsi que le nombre de valeurs propres sera de 14. Voici ci-dessous le tableau des valeurs propres ainsi que l'éboulis de celles-ci.

```
acm1$eig
```

```
##          eigenvalue percentage of variance cumulative percentage of variance
## dim 1  0.199871072      64.1631484           64.16315
## dim 2  0.026310111       8.4461424           72.60929
## dim 3  0.017947200       5.7614585           78.37075
## dim 4  0.013149413       4.2212599           82.59201
## dim 5  0.012693282       4.0748315           86.66684
## dim 6  0.009505299       3.0514167           89.71826
## dim 7  0.007811017       2.5075138           92.22577
## dim 8  0.006732469       2.1612753           94.38705
## dim 9  0.004917665       1.5786819           95.96573
## dim 10 0.003656827       1.1739243           97.13965
## dim 11 0.003044097       0.9772241           98.11688
## dim 12 0.002419275       0.7766423           98.89352
## dim 13 0.002232367       0.7166405           99.61016
## dim 14 0.001214371       0.3898406           100.00000
```

```
library(factoextra)
```

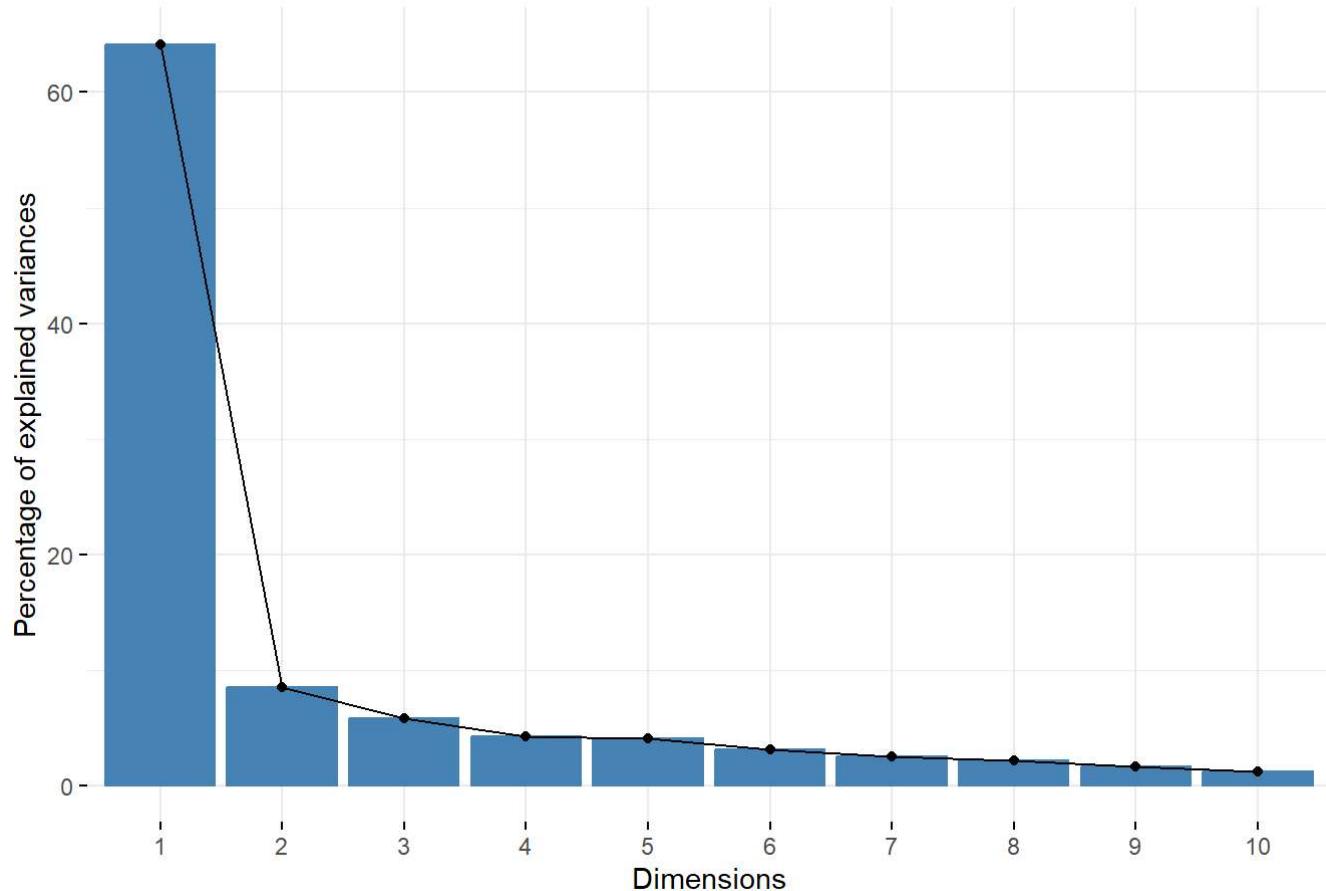
```
## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa
```

```
acm1$eig
```

```
##          eigenvalue percentage of variance cumulative percentage of variance
## dim 1  0.199871072      64.1631484           64.16315
## dim 2  0.026310111       8.4461424           72.60929
## dim 3  0.017947200       5.7614585           78.37075
## dim 4  0.013149413       4.2212599           82.59201
## dim 5  0.012693282       4.0748315           86.66684
## dim 6  0.009505299       3.0514167           89.71826
## dim 7  0.007811017       2.5075138           92.22577
## dim 8  0.006732469       2.1612753           94.38705
## dim 9  0.004917665       1.5786819           95.96573
## dim 10 0.003656827       1.1739243           97.13965
## dim 11 0.003044097       0.9772241           98.11688
## dim 12 0.002419275       0.7766423           98.89352
## dim 13 0.002232367       0.7166405           99.61016
## dim 14 0.001214371       0.3898406           100.00000
```

```
fviz_eig(acm1)
```

Scree plot



Si on utilise le critère du Coude pour choisir le nombre d'axe à retenir, nous pouvons décider de retenir 1 axe factoriel. En effet, la courbe se stabilise à partir de la deuxième valeur propre. Les deux premiers axes expliquent la majorité de l'inertie, on peut retenir les axes 1 et 2. Afin d'utiliser le critère de Kaiser, nous devons regarder les valeurs propres supérieures à la moyenne. Ici, l'inertie moyenne est de : $1/9 = 0,11$. Donc, ici, nous retiendrons qu'un seul axe, à savoir l'axe 1 (car $\lambda_1 = 0.19 > 1/9$).

Taux d'inertie modifiée:

Souvent en ACM, il est un peu délicat de choisir les axes par les deux critères utilisés ci-dessus. Dans ce cas on fait appel au Taux d'inertie modifiée. Enfin, utilisons le critère des taux d'inertie modifiés (en supposant que nous avons choisi trois axes): On calcule les deltas et les taux modifiés et on compare à $1/3=0.33$ Inertie totale $I = \frac{P}{S} - 1$ Le taux d'inertie modifiée associée à l'axe factoriel i est donné par:

$$\frac{\lambda'_i}{\sum_{j; \lambda_j > \frac{1}{s}} \lambda'_i}$$

avec λ_i les valeurs propres initiales et $\lambda'_i = ()^2 - \lambda_i^2$

calcul des taux d'inertie modifiée

```
s=9

delta=(acm1$eig[1:3,1]-1/s)^2

Iprime=delta/sum(delta);
Iprime
```

```
##      dim 1      dim 2      dim 3
## 0.3317324 0.3027998 0.3654678
```

```
s<-9
im<-1/s

lambda1<-acm1$eig[1,1]
lambda_1<-((s/(s-1))^2)*((lambda1 - (1/s))^2)

lambda2<-acm1$eig[2,1]
lambda_2<-((s/(s-1))^2)*((lambda2 - (1/s))^2)

lambda3<-acm1$eig[3,1]
lambda_3<-((s/(s-1))^2)*((lambda3 - (1/s))^2)

LambdaSum <- sum(lambda_1,lambda_2,lambda_3)
(lambdamodif1<-lambda_1/LambdaSum)
```

```
## [1] 0.3317324
```

```
(lambdamodif2<-lambda_2/LambdaSum)
```

```
## [1] 0.3027998
```

```
(lambdamodif3<-lambda_3/LambdaSum)
```

```
## [1] 0.3654678
```

Avec ce critère, on choisirais deux axes,l'axe 1 et l'axe 3 car $\lambda'_1 = 0.331$ et $\lambda'_3 = 0.36$. Par la suite , nous regarderons le nuage sur le plan (1,2) et éventuellement sur le plan(1,3).

Interprétation des résultats:

Contributions des modalités:

Nous allons, dans cette partie, analyser la contribution relative des différentes modalités sur les deux premiers axes. Pour cela, nous allons observer le tableau des contributions relatives des modalités ainsi que les diagrammes. La contribution moyenne d'une modalité sur un axe vaut $1/23 = 4,3\%$. Nous considérerons qu'une modalité contribue significativement à la formation de l'axe si sa contribution dépasse 4,3%.

```
acm1$var$contrib[,1:2]
```

```

##                               Dim 1      Dim 2
## urbain                  7.6684883  0.003065227
## rural                   5.3011556  0.002118964
## Eau en réservoir          6.2347414  7.823560229
## Source canalisée          2.3151935  7.277232726
## Source non protégée       2.4256795  0.487312593
## Source protégée           3.0567862  0.023215372
## chasse d'eau               6.4373594  10.970087455
## autre toilette             0.7348992  0.034301697
## latrine                   0.1980487  10.930647736
## aucune installation        5.1806121  5.805824206
## Tlv_non                   6.9637961  0.279669184
## Tlv_oui                   6.7680774  0.271809035
## Ref_non                   3.4533928  0.556272764
## Ref_oui                   9.7471861  1.570077431
## Lit_non                   5.1472521  8.127311783
## Lit_oui                   1.7891006  2.824920638
## Tab_non                   3.6657045  10.055411404
## Tab_oui                   1.0270014  2.817172353
## elec_gaz                  7.6507307  13.502859300
## charbon                   2.7915672  8.692369641
## bois                       6.4397781  0.032979614
## Net_non                   0.4370829  0.691144057
## Net_oui                   4.5663662  7.220636590

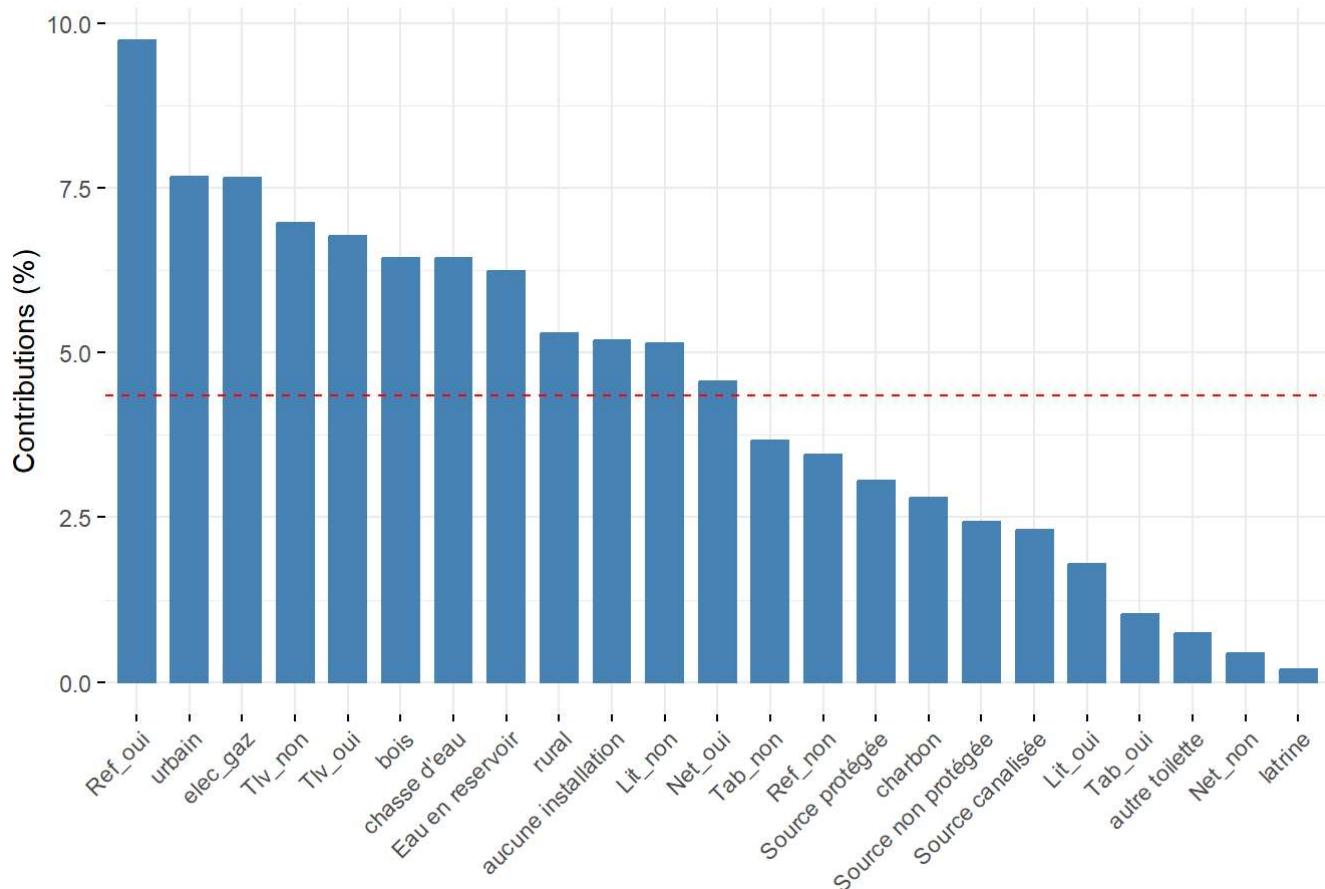
```

```

#acm1$var$coord[,1:2]
fviz_contrib(acm1, choice = "var", axes = 1)

```

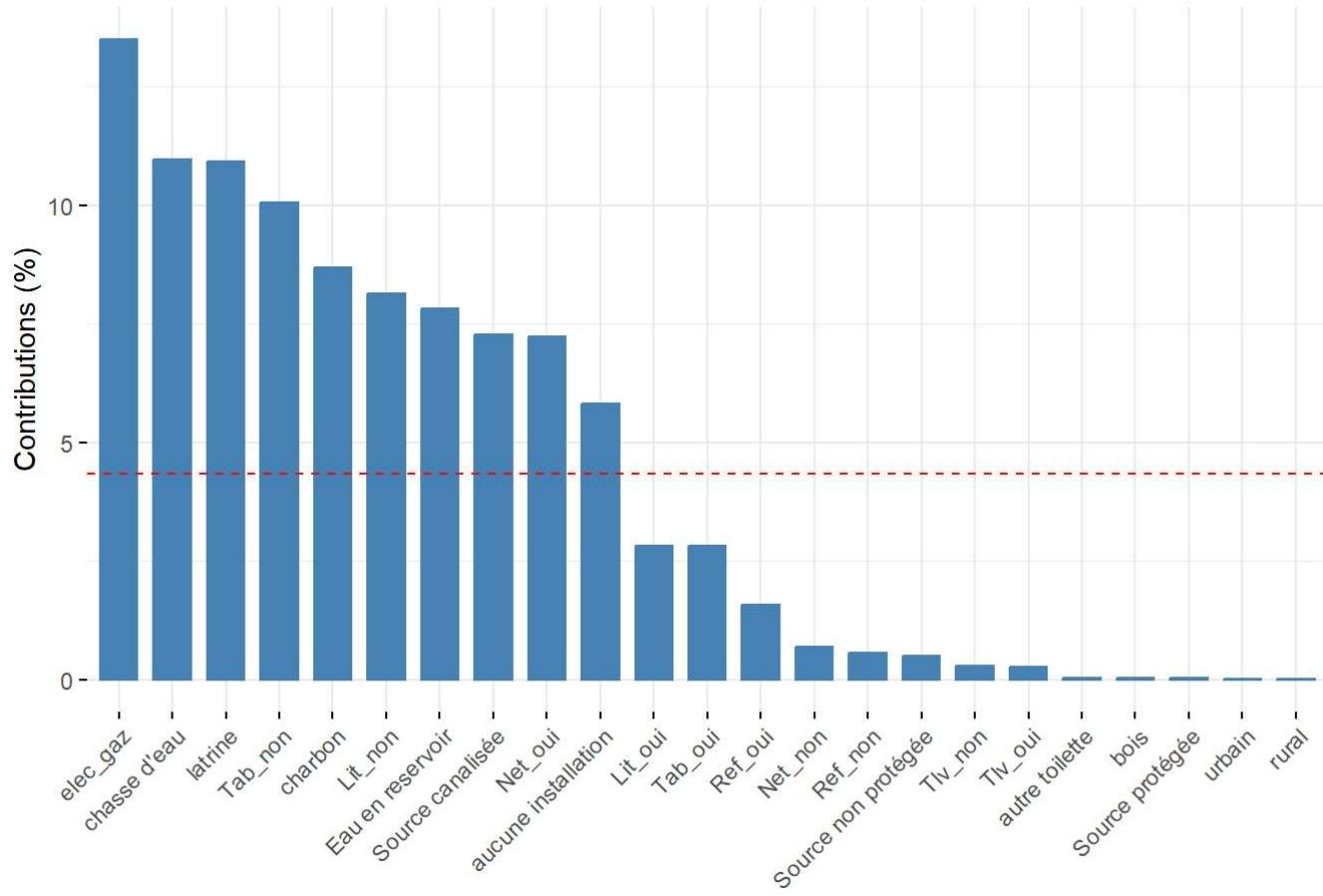
Contribution of variables to Dim-1



Pour le premier axe, nous observons que les modalités qui contribuent le plus à la formation de cet axe sont les suivantes : "Ref_oui" avec une contribution de 9.74%, "Urbain" avec 7.66%, "Elec_gaz" avec 7.65%, "_Tlv_non" avec 6.96%, "Tlv_oui" avec 6.76%, "Bois" avec 6.43%, "Chasse d'eau" avec 6.43%, "Eau en réservoir" avec 6.23%, "Rural" avec 5.30%, "Aucune installation" avec 5.18%, "Lit_non" avec 5.14%, et "Net_oui" avec 4.56%.

```
fviz_contrib(acm1, choice = "var", axes = 2)
```

Contribution of variables to Dim-2

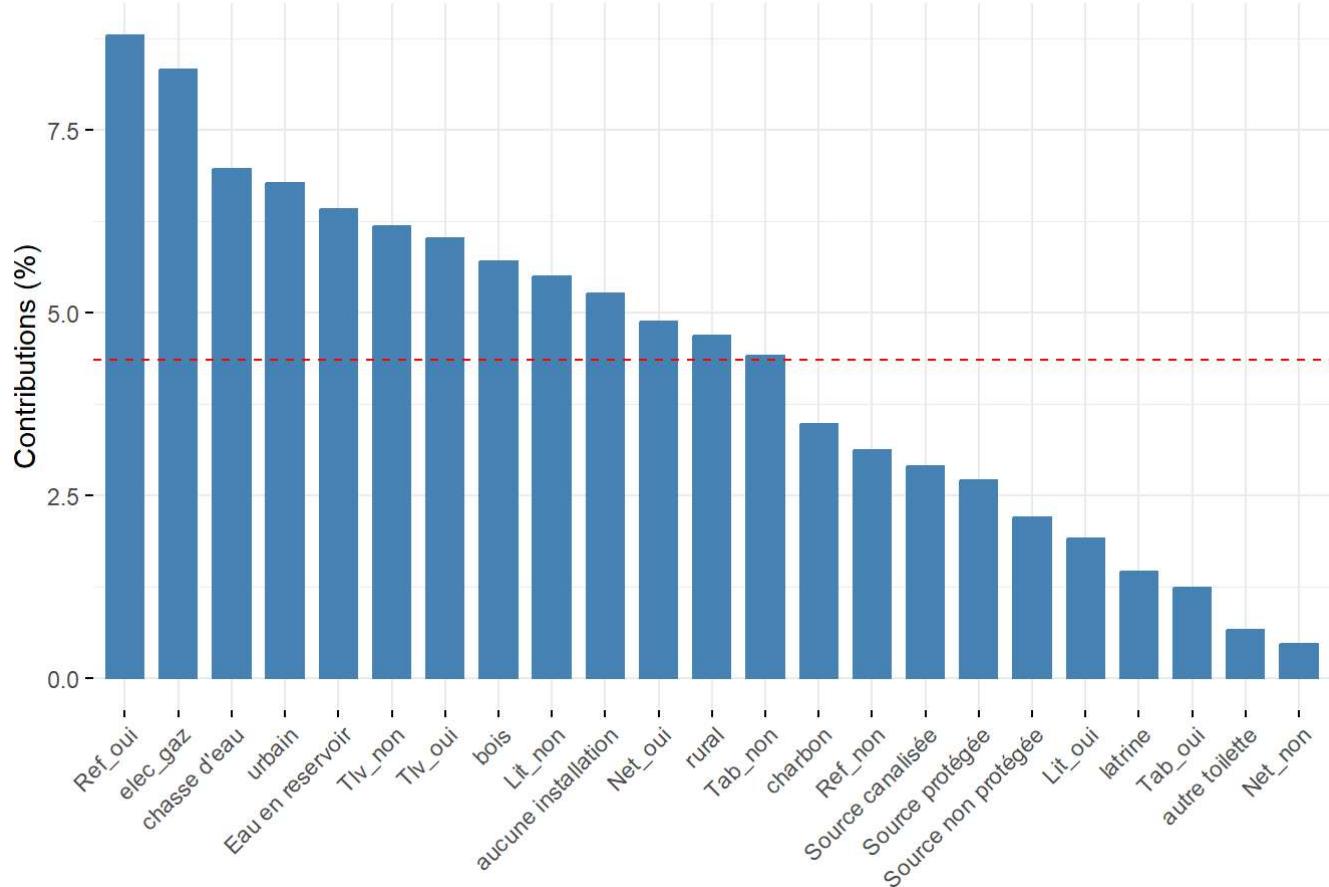


Pour le deuxième axe, les modalités qui présentent les contributions les plus significatives à la formation de cet axe sont les suivantes : "Elec_gaz" avec une contribution de 13.5%, "Chasse d'eau" avec 10.97%, "Latrine" avec 10.93%, "Tab_non" avec 10.05%, "Charbon" avec 8.69%, "Lit_non" avec 8.12%, "Eau de réservoir" avec 7.82%, "Source canalisée" avec 7.27%, "Net_oui" avec 7.22%, et "Aucune installation" avec 5.80%.

Ensuite, nous avons affiché graphiquement les modalités sur le premier plan factoriel en les distinguant par des couleurs en fonction de leur contribution à l'inertie totale. Nous ne procéderons pas à l'interprétation des modalités ayant une contribution minimale à la construction du plan factoriel.

```
fviz_contrib(acm1, choice = "var", axes = c(1,2))
```

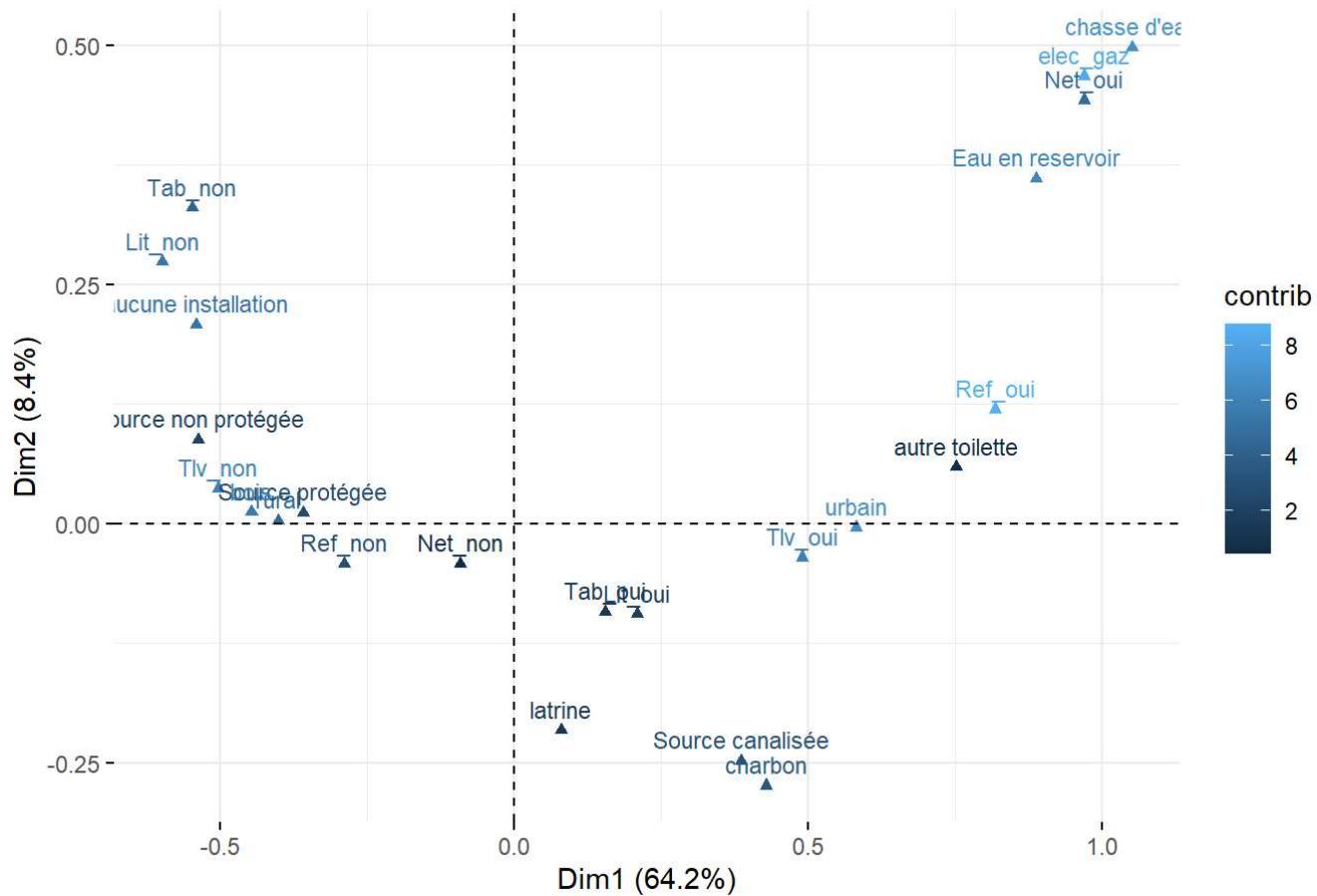
Contribution of variables to Dim-1-2



En observant les graphiques précédents, on constate que les modalités qui contribuent le plus significativement au plan sont Ref_oui, elec_gaz, chasse d'eau, urbain, Eau en réservoir, Tlv_non, Tlv_oui, bois, Lit_non, aucune installation, Net_oui, rural, Tab_non.

```
fviz_mca_var(acm1, col.var = "contrib",labelsize=3)
```

Variable categories - MCA



Qualité de représentations:

Ensuite, nous examinons les qualités de représentation des modalités sur les différents axes. Nous concentrons notre attention uniquement sur la qualité de représentation sur les deux premiers axes, car ils expliquent la majeure partie de l'inertie totale. Une qualité de représentation plus proche de 1 indique une représentation efficace de la modalité. Ainsi, considérerons qu'une modalité est bien représentée si sa qualité de représentation est supérieure à 0,50. Connaître, la qualité de représentation d'une modalité sur un axe est important avant de pouvoir l'interpréter.

En effectuant l'Analyse des Correspondances Multiples (ACM) avec R, nous obtenons le tableau suivant.

```
acm1$var$cos2[,1:2]
```

```

##                               Dim 1           Dim 2
## urbain                  0.86887966 0.0000457178
## rural                   0.86887966 0.0000457178
## Eau en réservoir         0.64779479 0.1070030538
## Source canalisée        0.37436860 0.1548998912
## Source non protégée     0.35086280 0.0092786388
## Source protégée          0.54525838 0.0005451119
## chasse d'eau              0.64656673 0.1450402093
## autre toilette            0.11220387 0.0006893954
## latrine                  0.06409562 0.4656665759
## aucune installation      0.67641477 0.0997857723
## Tlv_non                  0.89908287 0.0047530351
## Tlv_oui                  0.89908287 0.0047530351
## Ref_non                  0.88426128 0.0187497357
## Ref_oui                  0.88426128 0.0187497357
## Lit_non                  0.61818923 0.1284889108
## Lit_oui                  0.61818923 0.1284889108
## Tab_non                  0.48024539 0.1734116744
## Tab_oui                  0.48024539 0.1734116744
## elec_gaz                 0.70283593 0.1632862110
## charbon                  0.41326670 0.1693921331
## bois                      0.90882577 0.0006126720
## Net_non                  0.53098224 0.1105242246
## Net_oui                  0.53098224 0.1105242246

```

```
sort(acm1$var$cos2[,2],decreasing = TRUE)
```

##	latrine	Tab_oui	Tab_non	charbon
##	0.4656665759	0.1734116744	0.1734116744	0.1693921331
##	elec_gaz	Source canalisée	chasse d'eau	Lit_oui
##	0.1632862110	0.1548998912	0.1450402093	0.1284889108
##	Lit_non	Net_oui	Net_non	Eau en réservoir
##	0.1284889108	0.1105242246	0.1105242246	0.1070030538
##	aucune installation	Ref_non	Ref_oui	Source non protégée
##	0.0997857723	0.0187497357	0.0187497357	0.0092786388
##	Tlv_non	Tlv_oui	autre toilette	bois
##	0.0047530351	0.0047530351	0.0006893954	0.0006126720
##	Source protégée	rural	urbain	
##	0.0005451119	0.0000457178	0.0000457178	

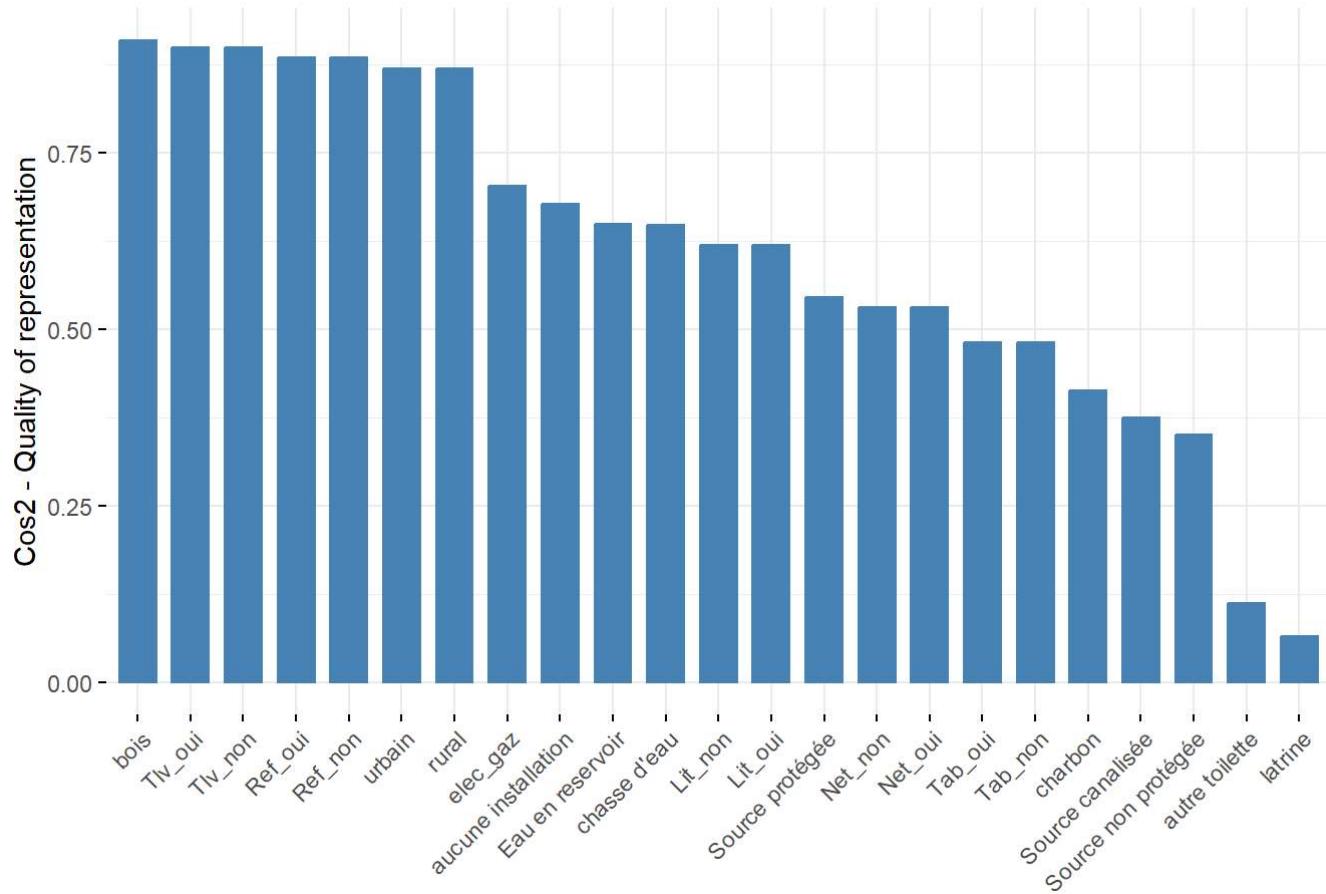
```
QLT12=acm1$var$cos2[,1]+acm1$var$cos2[,1]
```

```
QLT12
```

##	urbain	rural	Eau en réservoir	Source canalisée
##	1.7377593	1.7377593	1.2955896	0.7487372
## Source non protégée	Source protégée	chasse d'eau	autre toilette	
##	0.7017256	1.0905168	1.2931335	0.2244077
##	latrine	aucune installation	Tlv_non	Tlv_oui
##	0.1281912	1.3528295	1.7981657	1.7981657
##	Ref_non	Ref_oui	Lit_non	Lit_oui
##	1.7685226	1.7685226	1.2363785	1.2363785
##	Tab_non	Tab_oui	elec_gaz	charbon
##	0.9604908	0.9604908	1.4056719	0.8265334
##	bois	Net_non	Net_oui	
##	1.8176515	1.0619645	1.0619645	

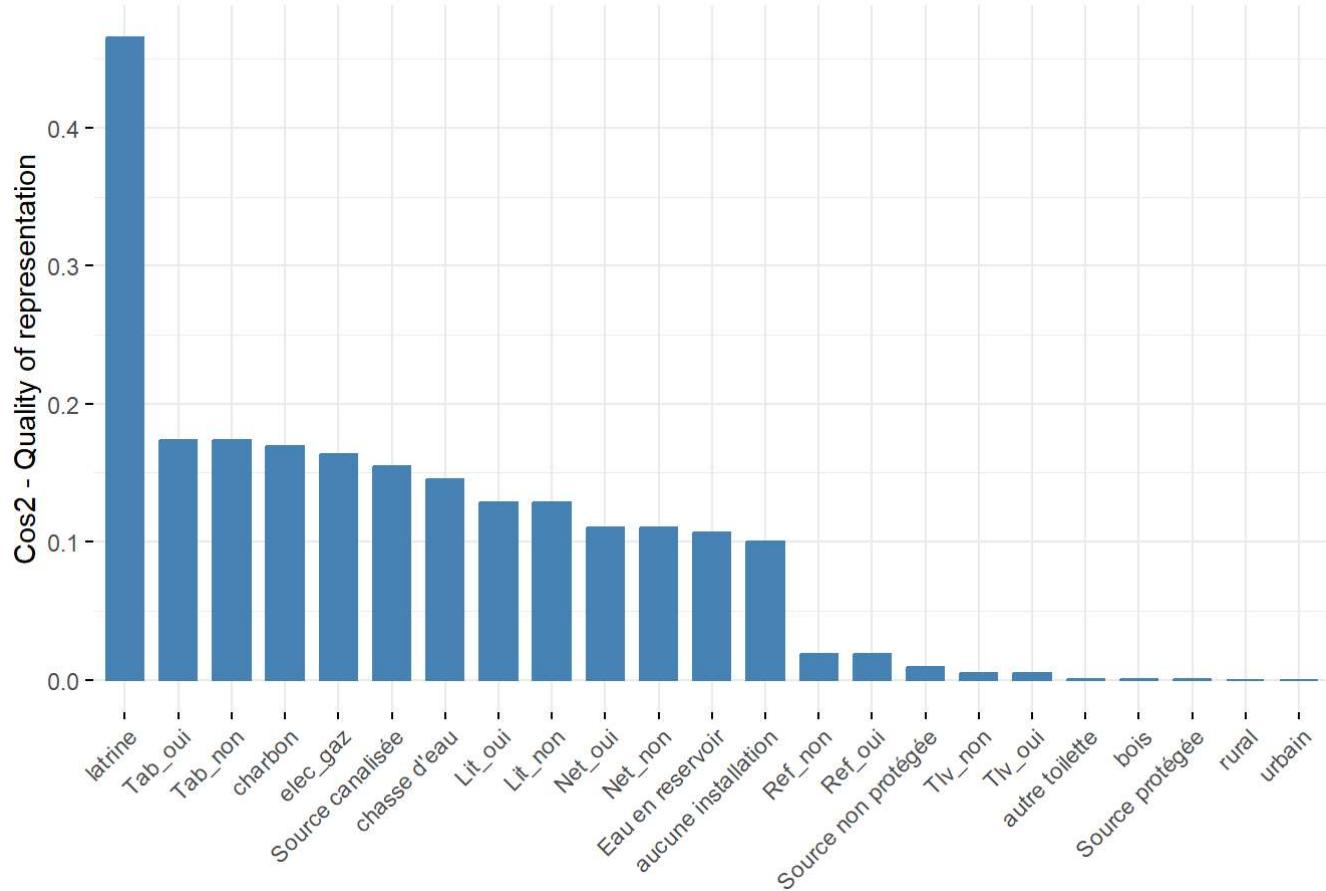
```
fviz_cos2(acm1, choice = "var", axes = 1)
```

Cos2 of variables to Dim-1



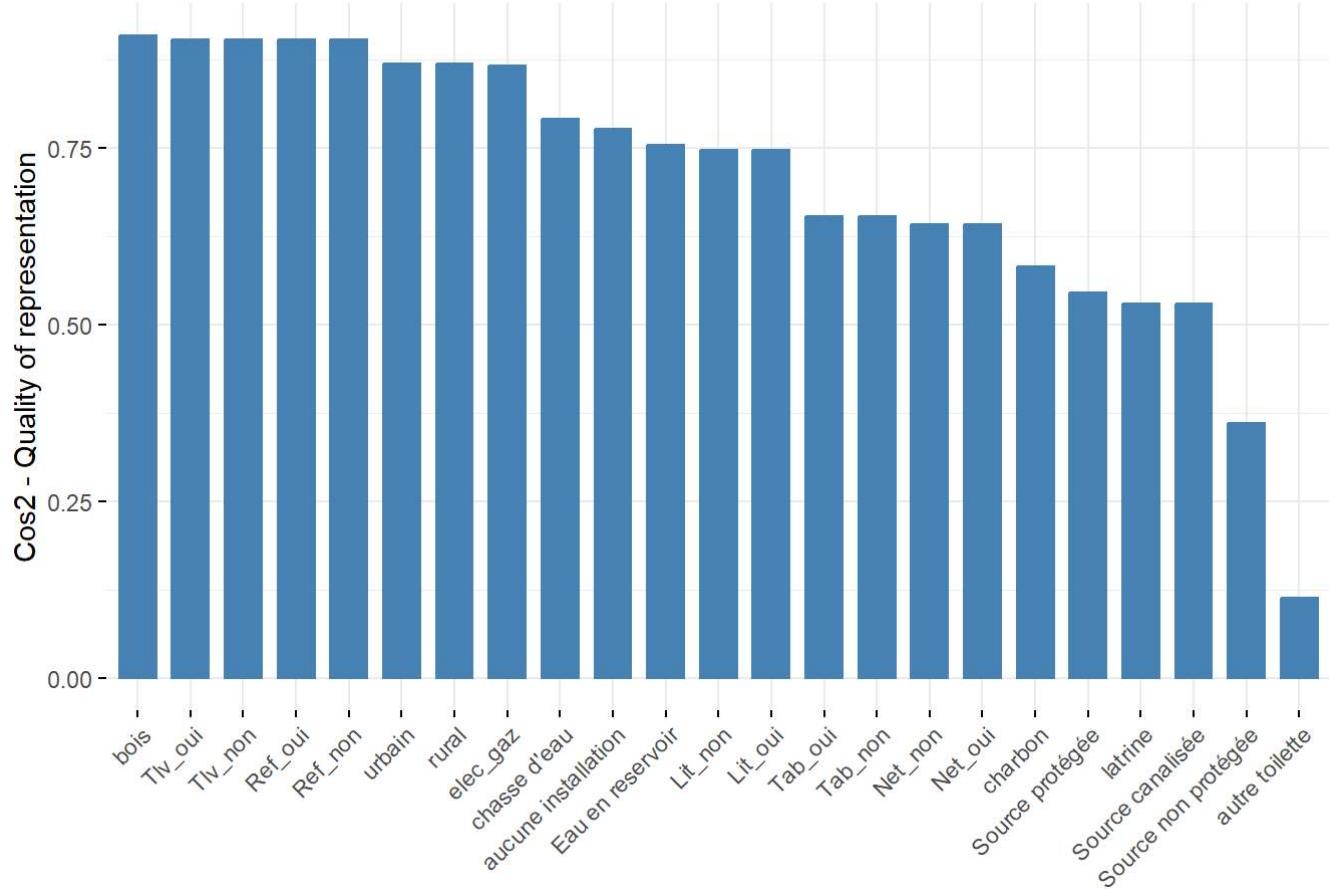
```
fviz_cos2(acm1, choice = "var", axes = 2)
```

Cos2 of variables to Dim-2



```
fviz_cos2(acm1, choice = "var", axes = 1:2)
```

Cos2 of variables to Dim-1-2



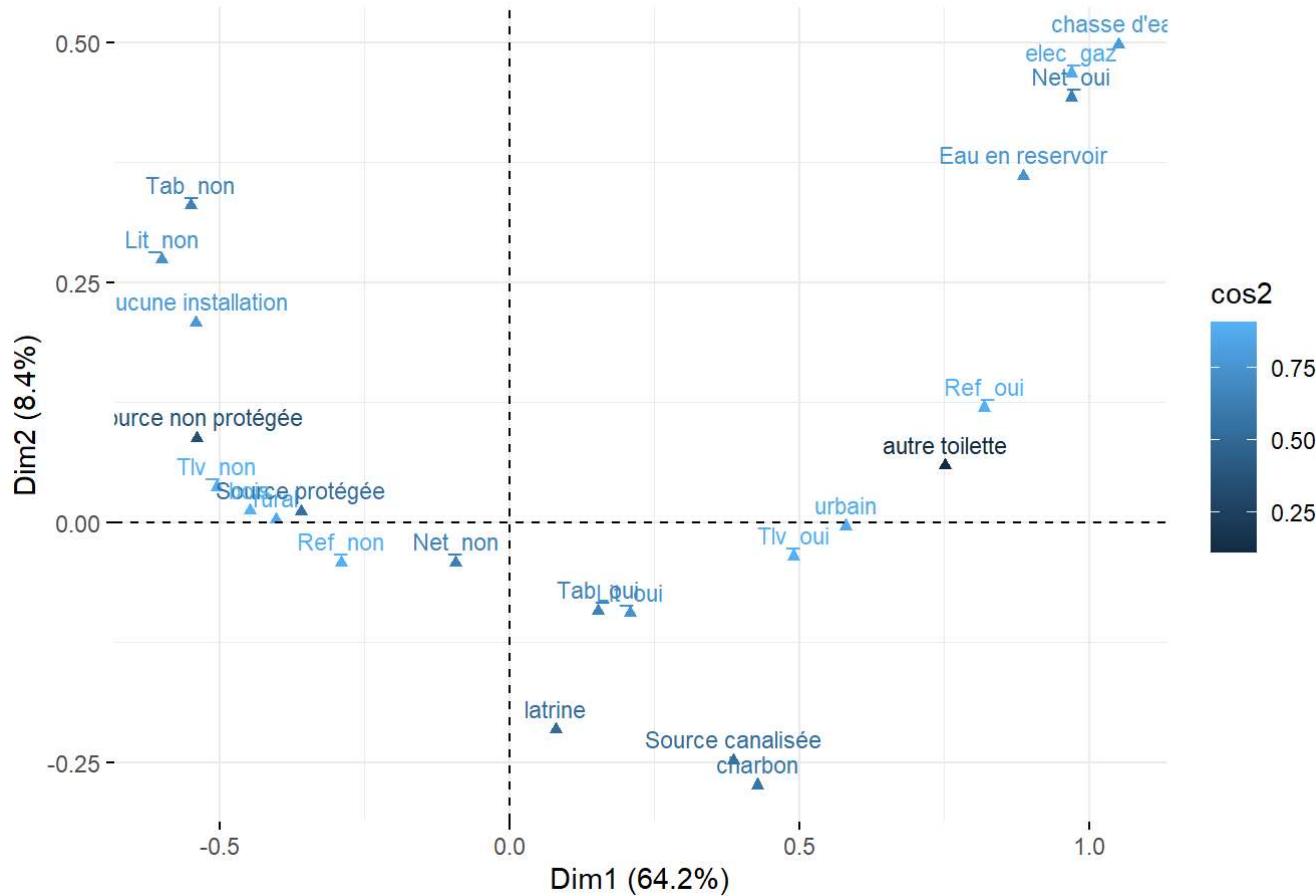
Ainsi, sur l'axe 1, les modalités telles que bois (0,90), Tlv_oui (0,89), Tlv_non (0,89), Ref_oui (0,88), Ref_non (0,88), urbain (0,86), rural (0,86), elec_gaz (0,70) sont considérées comme bien représentées. En revanche, les modalités aucune installation (0,67), Eau en réservoir (0,64), chasse d'eau (0,64), Lit_non (0,61), Lit_oui (0,61), Source protégée (0,54), Net_non (0,53), Net_oui (0,53) montrent une représentation moyennement satisfaisante.

Par contre, d'autres modalités, en particulier "autre toilette" (0,11) et "latrine" (0,06), affichent des qualités de représentation faibles.

Pour l'axe 2, la modalité ayant la plus grande qualité de représentation est latrine (0,47). Les autres modalités ont des qualités faibles.

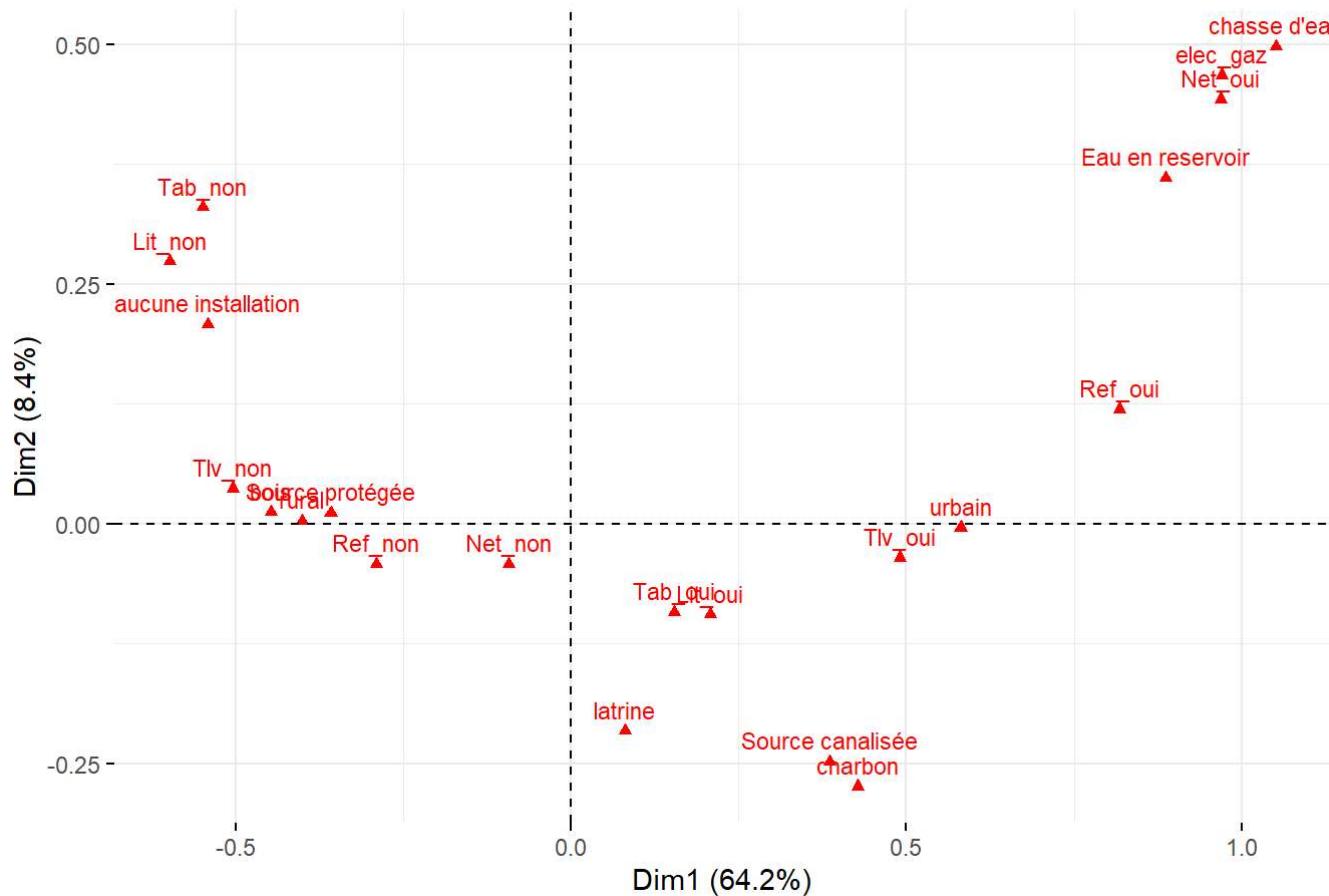
```
graphacm1 <- fviz_mca_var(acm1, col.var = "cos2", labelsize=3)
graphacm11 <- fviz_mca_var(acm1, select.var = list(cos2=0.5),labelsize=3)
graphacm1
```

Variable categories - MCA



graphacm11

Variable categories - MCA



```
#ggsave(graphacm1, file="graphacm1.pdf")
```

Interprétation des axes :

Axe 1 : À droite se trouvent les enfants de moins de 5 ans issus de ménages situés dans des zones urbaines, cuisinant avec du charbon, équipés d'une télévision, d'un lit, d'une table et d'une connexion internet, utilisant de l'eau potable provenant de sources canalisées, et disposant de latrines pour les toilettes. En revanche, à gauche se situent les enfants vivant en zones rurales, avec des mères utilisant du bois pour cuisiner, dans des ménages dépourvus d'installations sanitaires, et ne possédant ni télévision, ni lit, ni table.

L'axe 1 met en opposition les enfants résidant dans des ménages plus aisés en zones urbaines et ceux résidant dans des ménages moins favorisés en zones rurales.

Axe 2 : L'axe 2 distingue les enfants provenant de ménages qui cuisinent au gaz ou à l'électricité, utilisent de l'eau de réservoir comme source d'eau potable, et disposent de toilettes avec chasse d'eau. En revanche, il différencie ceux vivant dans des ménages cuisinant avec du charbon et ayant des latrines comme installation sanitaire. Cet axe met en relief les différences entre les enfants de ménages aisés et ceux de ménages plus modestes.

Interpretation de la proximité entre les modalités:

Deux modalités sont proches signifie que ce sont les mêmes individus ou des individus semblables qui les possèdent.

Les modalités "Chasse d'eau", "Elec_gaz", "Net_oui", et "Eau réservoir" présentent des similitudes dans leurs profils. Cette observation suggère que les ménages où l'on cuisine avec du gaz ou de l'électricité et qui disposent de toilettes à chasse d'eau sont vraisemblablement également équipés d'internet et utilisent de l'eau provenant de citernes ou de bouteilles comme source d'eau potable.

Les modalités "Tab_non", "Lit_non", et "Aucune installation" montrent des profils similaires, suggérant que les ménages où les enfants ne disposent d'aucune installation pour les toilettes ont probablement une forte probabilité de ne pas avoir de lit ni de table.

Inversement, les modalités "Urbain", "Tlv_oui", et "Ref_oui", ainsi que les modalités "Rural", "Tlv_non", et "Ref_non", présentent des profils opposés, ce qui semble cohérent étant donné les différences entre les ménages ruraux et urbains.

ACM du tableau disjonctif

```
disj = acm.disjonctif(menage2)
head(disj)
```

```

## Zn.urbain Zn.rural EP.Eau en réservoir EP.Source canalisée
## 1      0      1      0      0
## 2      0      1      0      0
## 3      0      1      0      0
## 4      0      1      0      0
## 5      0      1      0      0
## 6      0      1      0      0
## EP.Source non protégée EP.Source protégée Toil.chasse d'eau
## 1          0          1          0
## 2          0          1          0
## 3          0          1          0
## 4          0          1          0
## 5          0          1          0
## 6          0          1          0
## Toil.autre toilette Toil.latrine Toil.aucune installation Tlv.non Tlv.oui
## 1          0          1          0          1          0
## 2          0          0          1          0          1
## 3          0          1          0          0          1
## 4          0          1          0          0          1
## 5          0          1          0          1          0
## 6          0          1          0          0          1
## Ref.non Ref.oui Lit.non Lit.oui Tab.non Tab.oui Comb.elec_gaz Comb.charbon
## 1      1      0      0      1      0      1      0      0
## 2      1      0      0      1      0      1      0      1
## 3      0      1      1      0      1      0      0      1
## 4      0      1      0      1      0      1      0      0
## 5      1      0      0      1      0      1      0      0
## 6      1      0      0      1      0      1      0      1
## Comb.bois Net.non Net.oui
## 1      1      1      0
## 2      0      1      0
## 3      0      0      1
## 4      1      0      1
## 5      1      1      0
## 6      0      1      0

```

```

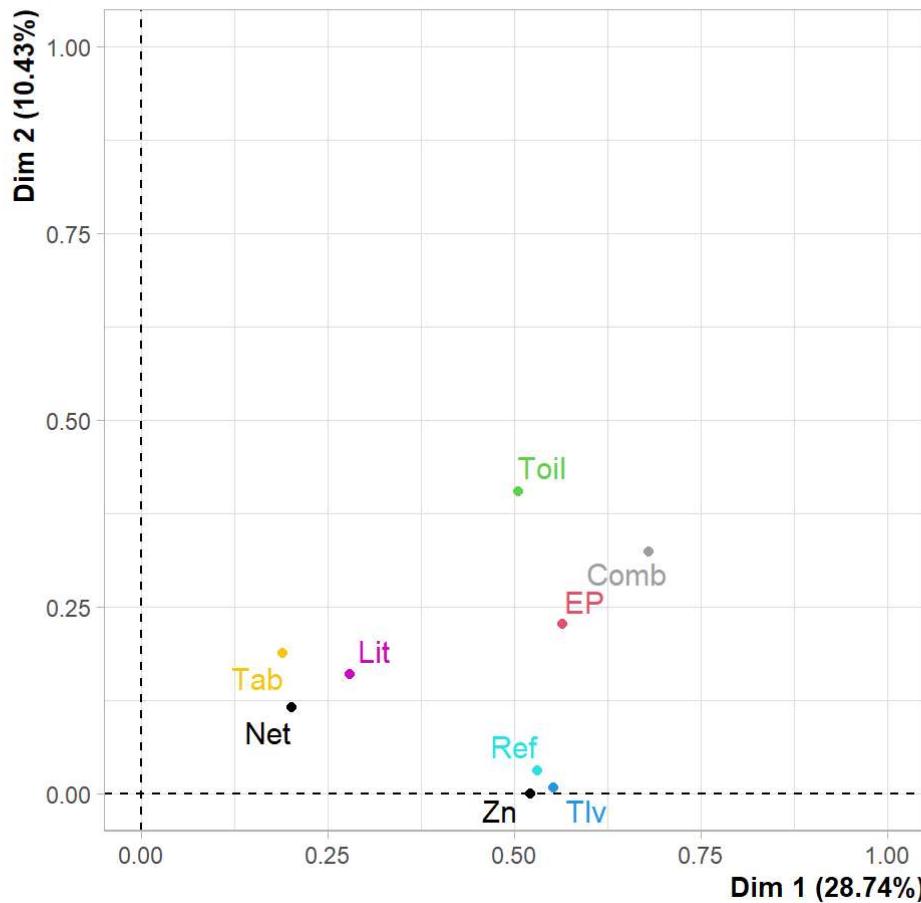
library(factoextra)
library(FactoMineR)
res.MCA1<-MCA(menage2,graph=F)

```

```

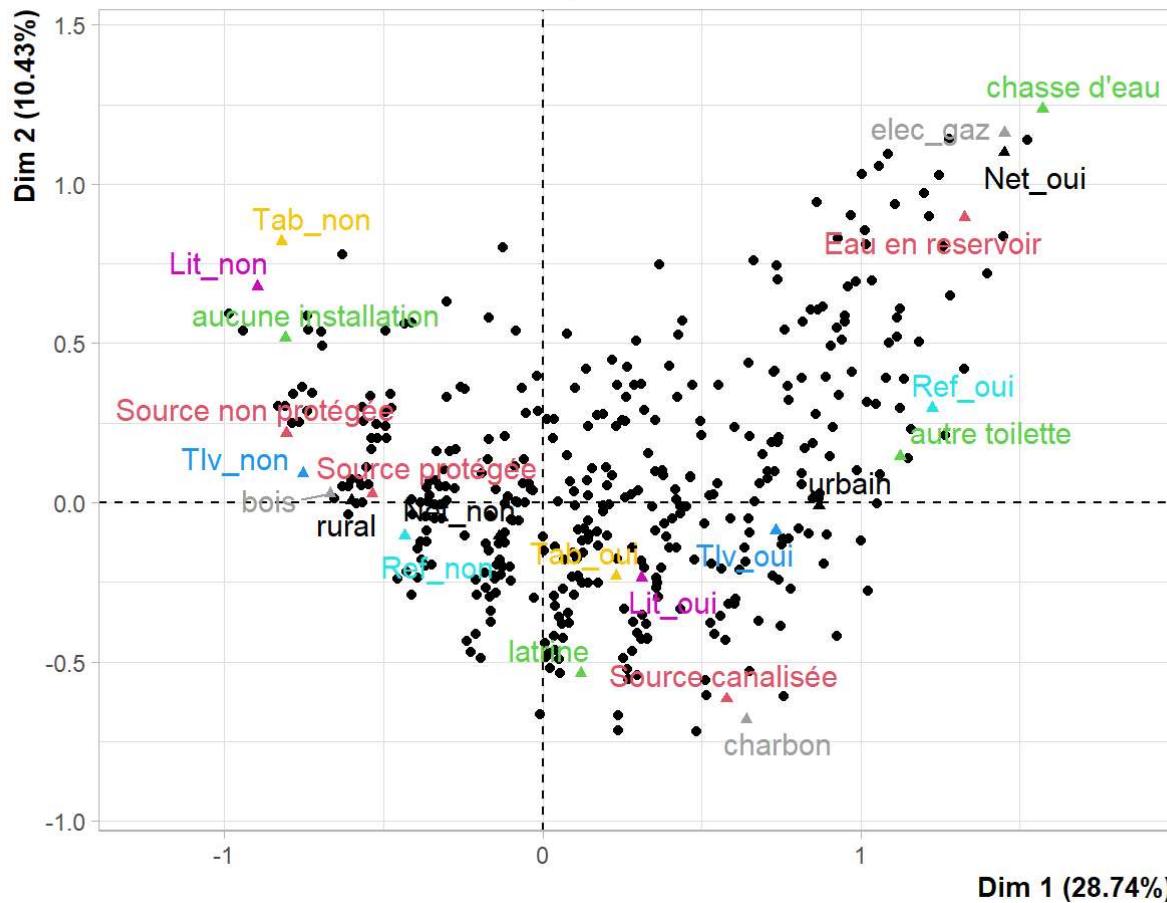
res.MCA<-MCA(menage2,graph=FALSE)
plot.MCA(res.MCA, choix='var',title="Graphe des variables",col.var=c(1,2,3,4,5,6,7,8,9))

```

Graphe des variables

```
graphacm2 <- plot.MCA(res.MCA,col.var=c(1,1,2,2,2,2,3,3,3,3,4,4,5,5,6,6,7,7,8,8,8,9,9),title="Graphe de l'ACM",label =c('var'))
graphacm2
```

Graphe de l'ACM



```
#ggsave(graphacm2, file="graphacm2.pdf")
```

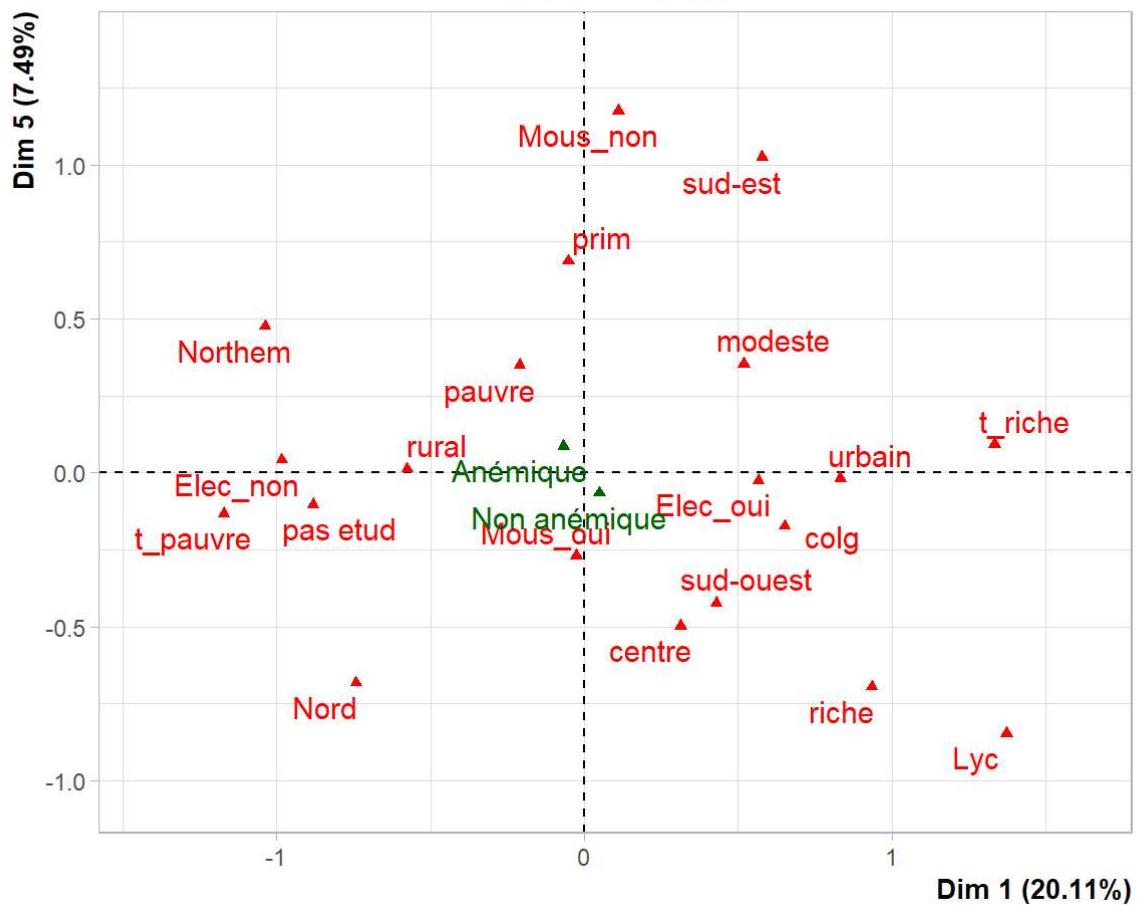
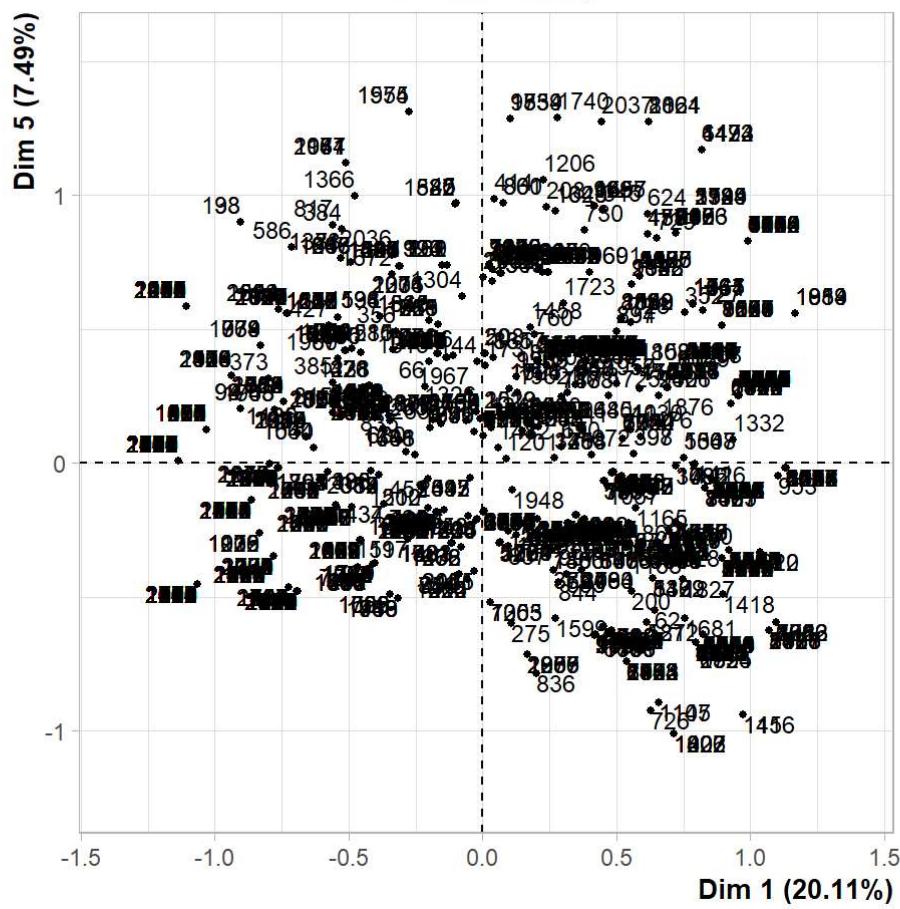
Selon l'observation de la figure ci-dessus, on constate que les enfants positionnés dans la partie supérieure droite du quadrant résident au sein de ménages aisés, équipés d'internet, de cuisinières à gaz ou électriques, et disposant de toilettes à chasse d'eau. En revanche, la majorité des enfants semble concentrée au centre du quadrant, suggérant qu'ils habitent dans des foyers plus modestes caractérisés par la possession de tables et de lits, mais dépourvus d'accès à internet et dotés de toilettes équipées de latrines.

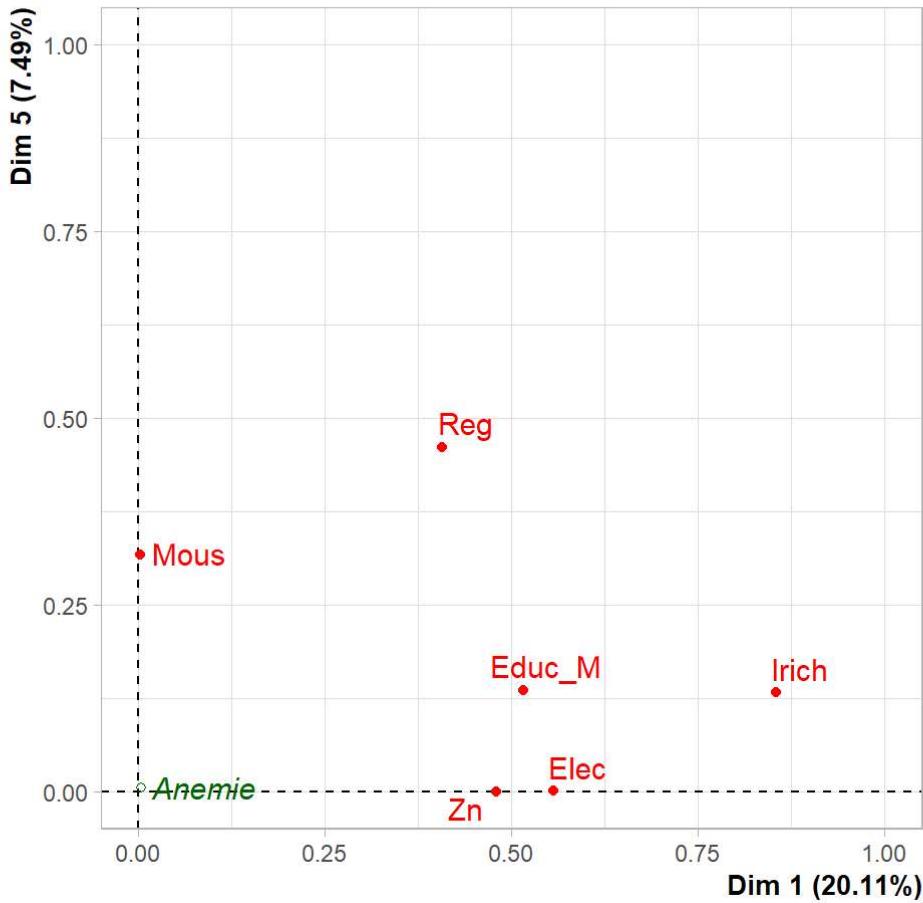
Une minorité d'enfants provient de ménages très défavorisés, dépourvus de lits, de tables et d'installations sanitaires. Ces ménages très pauvres se distinguent par une absence d'équipements de base.

ACM en ajoutant les variables supplémentaires et valeurs tests.

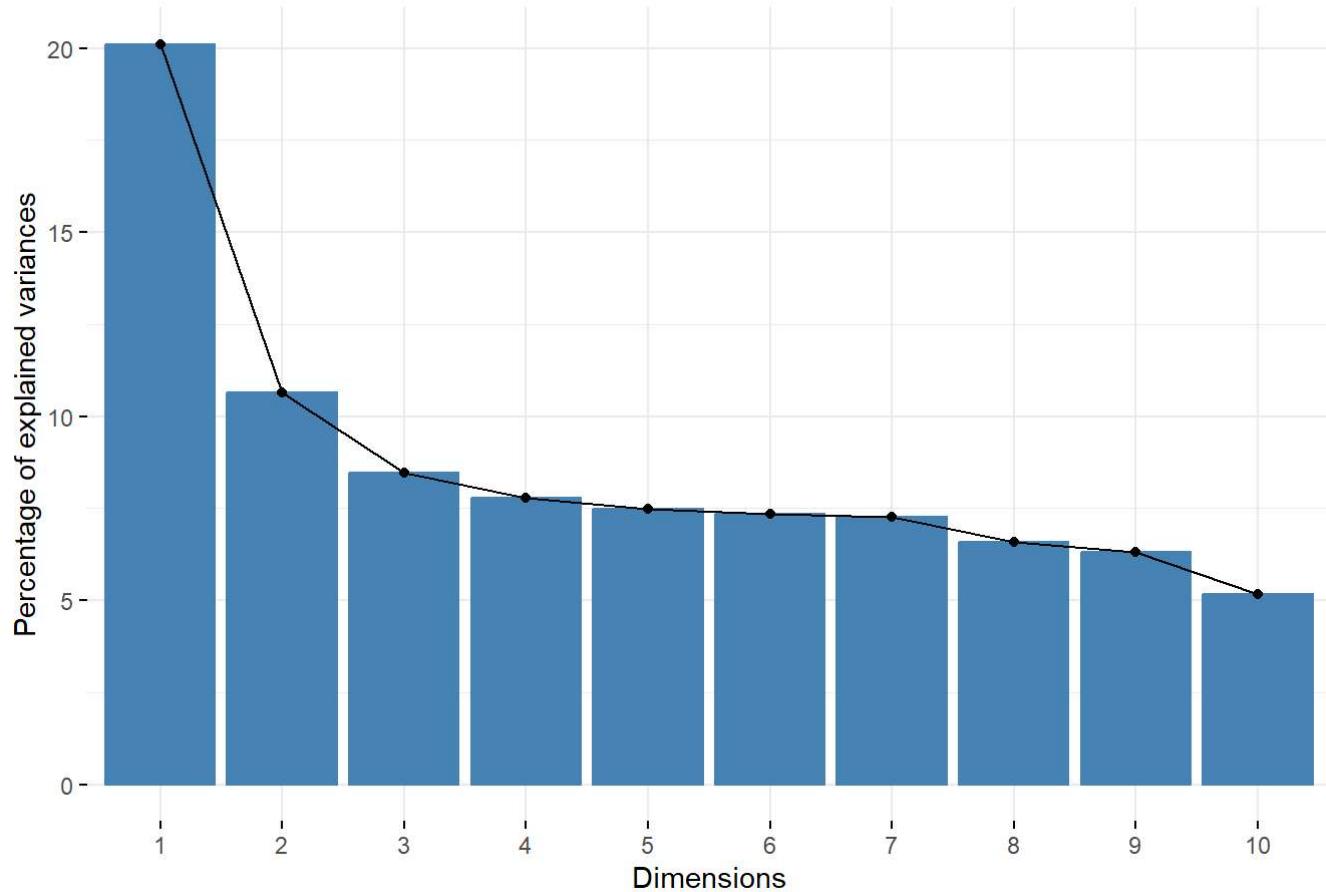
Nous avons décidé de faire l'ACM avec les variables concernant notre étude à savoir Niveau d'étude de la mère, la zone des enfants, les régions, la possession de moustiquaire, l'indice de richesse et l'électricité comme variables actives. Et en variable supplémentaire notre variable à expliquer l'Anémie. En effet, nous avons observé que si nous conservons les mêmes variables actives que précédemment et que nous rajoutons en supplémentaire les 5 nouvelles variables, l'ACM ne rendait pas un résultat exploitable pour notre problématique, du fait que ces variables concernaient plus les équipements des ménages alors que notre problématique est sanitaire.

```
menage3 <- menage1[,c(1:7)]
acm3 <- MCA(menage3, quali.sup = c(1), graph = TRUE, axes=c(1,5))
```

MCA factor map**MCA factor map**

Variables representation

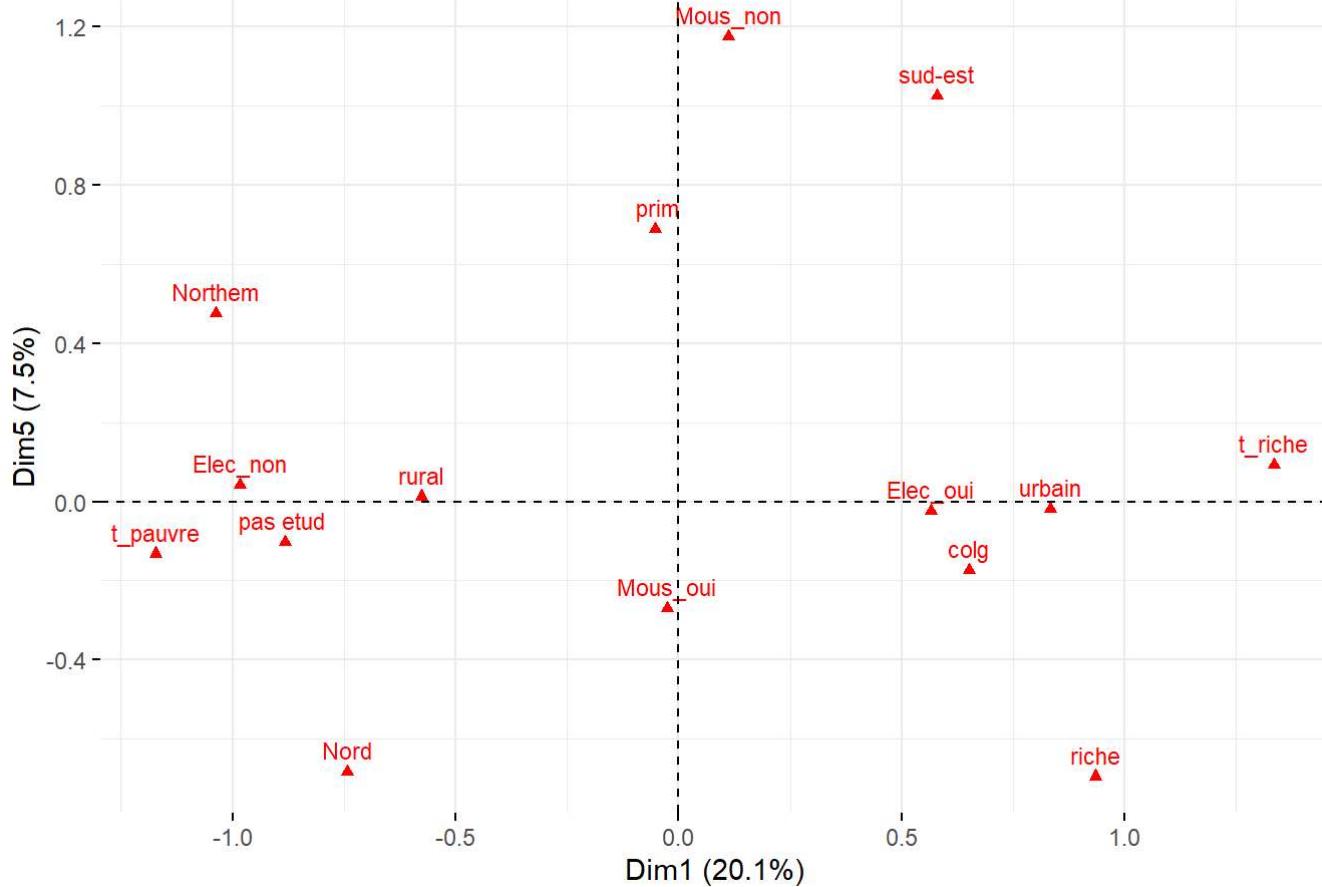
```
fviz_eig(acm3)
```

Scree plot

Si l'on utilisait le critère du coude on retiendrait que les deux premiers axes soit 30,7% de l'inertie totale. Mais pour des raisons d'interprétabilité, on a considérer le plan (1,5). Sur ce plan on conserve 27,6 % de l'information totale.

```
graphacm3 <- fviz_mca_var(acm3, col.var = "cos2", labelsize=3, axes = c(1,5))
fviz_mca_var(acm3, select.var = list(cos2=0.1),labelsize=3, axes=c(1,5))
```

Variable categories - MCA



```
#ggsave(graphacm3, file="graphacm3.pdf")
```

En analysant la distribution des modalités en relation avec la variable supplémentaire "Anémie", nous observons une tendance marquée. Du côté droit, un regroupement de modalités telles que "Elec_oui", "urbain", "colg", "t_riche", avec la modalité "Non_anémique" qui leur est proche. Cela suggère que les enfants issus de ménages aisés, bénéficiant de l'électricité, résidant dans des zones urbaines, et dont les mères ont atteint au moins un niveau collégial, ont de fortes chances de ne pas être anémiques.

À l'inverse, du côté gauche, les modalités "rural", "Elec_non", "pas_etud", "t_pauvre" se regroupent, et la modalité "anémique" est proche d'elles. Cette observation indique que la plupart des enfants anémiques proviennent de ménages pauvres, sans électricité, situés en milieu rural, et dont les mères n'ont pas suivi d'études.

Par ailleurs, les modalités "Anémique" et "Non_anémique", ainsi que "Mous_oui" et "Mous_non", sont opposées deux à deux le long de l'axe 2. Cela suggère que les enfants qui utilisent des moustiquaires autour de leurs lits sont moins susceptibles de souffrir de l'anémie que ceux qui n'en ont pas.

Ainsi, il semble que le niveau d'éducation de la mère, le niveau de richesse du ménage, la zone de résidence, la disponibilité d'électricité et la possession de moustiquaires ont un impact significatif sur la prévalence de l'anémie chez les enfants.

Il est notable que le nuage de modalités souligne davantage l'opposition entre les enfants provenant de ménages aisés en zones urbaines, dont les mères ont suivi des études, et ceux issus de ménages défavorisés en zones rurales, dont les mères n'ont pas poursuivi d'études, plutôt que l'opposition entre les enfants anémiques et non anémiques. Cette observation peut s'expliquer par les corrélations significatives existant entre le niveau d'éducation de la mère, la richesse du ménage et la zone de résidence ainsi que la taille de l'échantillon. Néanmoins, la relation avec l'anémie demeure pertinente.

Valeur test

```
acm3$quali.sup$v.test[,c(1:5)]
```

```
## Dim 1      Dim 2      Dim 3      Dim 4      Dim 5
## Anémique   -2.674547 -2.670043 -2.038518  2.89686  3.514379
## Non anémique 2.674547  2.670043  2.038518 -2.89686 -3.514379
```

Ce résultat prouve le choix de notre plan, car les contributions des modalités Anémie et Non_anémique sur l'axe 5 sont plus importantes que sur les autres .

```
acm3$quali.sup$coord[,c(1:5)]
```

```
## Dim 1      Dim 2      Dim 3      Dim 4      Dim 5
## Anémique   -0.06649359 -0.06638161 -0.05068087  0.07202065  0.08737316
## Non anémique 0.04948360  0.04940027  0.03771600 -0.05359676 -0.06502189
```

```
acm3$quali.sup$coord[,1:5]^2/sum(acm3$quali.sup$coord[,c(1:5)]^2)
```

```
## Dim 1      Dim 2      Dim 3      Dim 4      Dim 5
## Anémique   0.11749824 0.11710283 0.06825901 0.13784335 0.2028747
## Non anémique 0.06507204 0.06485305 0.03780272 0.07633942 0.1123546
```

```
acm3$quali.sup$coord[,5]^2/sum(acm3$quali.sup$coord[,c(1:5)]^2)
```

```
## Anémique Non anémique
## 0.2028747  0.1123546
```

On remarque que la qualité de représentation des modalités Anémie et Non_anémique sur l'axe 5 sont meilleures que sur les autres axes.

D'après la sortie R ci-dessus, nous pouvons formuler les commentaires suivants :

La modalité "Anémique" se positionne au-dessus de l'axe 1 et légèrement à gauche de l'axe 2. En revanche, la modalité "Non_anémique" se trouve en dessous de l'axe 1, juste à gauche de l'axe 2. Cela suggère que les enfants issus de ménages où les mères ont atteint un niveau d'éducation supérieur au primaire sont plus susceptibles d'être anémiques, tandis que ceux dont les mères ont fait des études et proviennent de ménages plus aisés sont positionnés davantage à droite de l'axe 2.

Résumé:

Les résultats des tris croisés ont révélé des constatations intéressantes. La représentation graphique de la prévalence de l'anémie en fonction de la richesse du ménage a mis en lumière que le pourcentage d'enfants anémiques est plus élevé dans les ménages pauvres ou modestes que dans les ménages aisés.

De plus, une corrélation significative a été observée entre le niveau d'éducation de la mère et la présence d'anémie chez les enfants. Une relation semble également exister entre l'anémie infantile et la région de résidence du ménage.

Dans les foyers dépourvus d'électricité, le taux d'enfants anémiques est plus élevé que dans les foyers avec accès à l'électricité. Il est à noter également que le taux de cas modérés est plus élevé dans les ménages sans électricité.

Enfin, une disparité de taux d'enfants anémiques a été observée entre les zones rurales et urbaines. Notamment, le pourcentage de cas modérés est plus élevé en milieu rural qu'en milieu urbain.

De manière similaire, l'analyse des Correspondances Multiples (ACM) que nous avons effectuée a pris en compte les variables actives telles que l'accès au niveau d'éducation de la mère, le niveau de richesse, la zone de résidence, la possession de moustiquaires, l'accès à l'électricité, la région, avec comme variable supplémentaire l'anémie chez les enfants.

En synthèse, l'analyse de la distribution des modalités par rapport à la variable supplémentaire "Anémie" révèle des tendances marquées. Les enfants issus de ménages aisés, ayant accès à l'électricité, résidant en milieu urbain, et dont les mères ont atteint au moins un niveau collégial, présentent une forte probabilité de ne pas être anémiques.

En revanche, la plupart des enfants anémiques proviennent de ménages pauvres, sans accès à l'électricité, vivant en milieu rural, et dont les mères n'ont pas suivi d'études.

Par ailleurs, les résultats indiquent une corrélation entre l'utilisation de moustiquaires et la prévalence de l'anémie. Les enfants utilisant des moustiquaires semblent moins susceptibles de souffrir d'anémie par rapport à ceux qui n'en ont pas. En résumé, le niveau d'éducation de la mère, le niveau de richesse du ménage, la zone de résidence, l'accès à l'électricité, et la possession de moustiquaires semblent exercer un impact significatif sur la fréquence de l'anémie chez les enfants. L'analyse univariée, bivariée et l'analyse des Correspondances Multiples (ACM) ont abouti à des résultats cohérents répondant à notre question de recherche : le niveau d'éducation maternelle, le niveau de richesse des ménages, le milieu de résidence des enfants et l'accès à l'électricité ont un impact sur le facteur de malnutrition, notamment l'anémie, chez les enfants au Ghana.