

**AGENCE NATIONALE DE LA STATISTIQUE ET DE LA
DEMOGRAPHIE**



**ECOLE NATIONALE DE LA STATISTIQUE ET DE L'ANALYSE
ECONOMIQUE PIERRE NDIAYE**

TP 2 - Logiciel statistique R

Par

KPAKOU M'Mounéné ISEP3

Chargé du cours:

Mr. Aboubacar HEMA

Ingénieur des Travaux Statistiques

Libraries

```
library(labelled)# labelisation
library(tidyverse)
library(readxl)# lecture des fichiers excels
library(sjPlot)#création de graphiques descriptifs
library(dplyr)
library(gtsummary)# tableau statistique
```

1. Importation et mise en forme

Repertoire de travail

```
setwd(getwd())
```

Importation

```
base_tp2 <- read_xlsx("Base TP2.xlsx")
#View(projet)
```

```
dim(base_tp2)
```

Nombre de lignes et colonnes

```
[1] 53 30
```

La base de données contient 53 lignes et 30 colonnes

Recodage et labelisation**Labelisation des variables**

On va ajouter des labels aux variables afin de les rendre plus explicites c'est à dire les labeliser

```
variables <- c("region","departement","sexe","age","sit_mat","si_chef_men","ethnie",
              "occupation","formation","niveau_alphabs","types_varietes",
              "types_varietes_1","types_varietes_2")

lab<- c("region","departement","sexe","age","situation matrimoniale",
        "Statut dans le menage","ethnie","occupation","formation",
        "niveau d'alphabétisation","type de variétés","Traditionnelles","Améliorées")
for (i in seq_along(variables)) {
  var_label(base_tp2[[variables[i]]]) <- lab[i]
}
```

Labelisation des variables commençant par criteres*

```
# Liste des noms des variables commençant par "criteres"
variables_critères <- grep("^criteres", names(base_tp2), value = TRUE)

var_types<- c("critères de choix ", "rendements élevés", "tailles de graines", "résistante",
              "tolerante aux sécheresses", "tolérantes aux inondations",
              "faible charge de travail", "faibles quantités d'intrants",
              "facile à transformer", "haute teneur en huile", "haut rendement après trans",
              "Demande sur le marché", "Bon goût", "Belle couleur",
              "Haut rendement en fourages", "Qualité du fourage",
              "Autre à spécifier")

# Attribution des étiquettes aux variables correspondantes
for (i in seq_along(variables_critères)) {
  var_label(base_tp2[[variables_critères[i]]]) <- var_types[i]
}
```

Vérifions pour voir si les étiquettes ont été bien appliquées à toutes les variables

```
# Exemple des vvariables commençant par "crit"
base_tp2 |>
  look_for("crit")
```

pos	variable	label	col_type	missing
14	criteres_var	critères de choix	chr	0
15	criteres_var_1	rendements élevés	dbl	0
16	criteres_var_2	tailles de graines	dbl	0
17	criteres_var_3	résistantes aux maladies/ ravageurs	dbl	0
18	criteres_var_4	tolerante aux sécheresses	dbl	0
19	criteres_var_5	tolérantes aux inondations	dbl	0
20	criteres_var_6	faible charge de travail	dbl	0
21	criteres_var_7	faibles quantités d'intrants	dbl	0
22	criteres_var_8	facile à transformer	dbl	0
23	criteres_var_9	haute teneur en huile	dbl	0
24	criteres_var_10	haut rendement après transformation	dbl	0
25	criteres_var_11	Demande sur le marché	dbl	0
26	criteres_var_12	Bon goût	dbl	0
27	criteres_var_13	Belle couleur	dbl	0
28	criteres_var_14	Haut rendement en fourages	dbl	0
29	criteres_var_15	Qualité du fourage	dbl	0
30	criteres_var_16	Autre à spécifier	dbl	0

values

Recodage des modalités

Nous allons utiliser la fonction **recode_factor** du package **dplyr**

- Sexe

```
base_tp2$sexe <- recode_factor(base_tp2$sexe,  
                               `1`="Homme", `2`= "Femme" )
```

- *situation matrimoniale*

```
base_tp2$sit_mat <- recode_factor(base_tp2$sit_mat,  
                                  `1` = "Marié(e)",  
                                  `3` = "Veuf(ve)",  
                                  `4` = "Divorcé(e)",  
                                  `5` = "Séparé(e)",  
                                  `6` = "Célibataire")
```

- *si_chef_menage*

```
base_tp2$si_chef_men <- recode_factor(base_tp2$si_chef_men,  
                                       `1` = "Femme du chef de ménage",  
                                       `2` = "Chef de ménage",  
                                       `3` = "Fils-fille du chef de ménage",  
                                       `99` = "Autres")
```

- *ethnie*

```
base_tp2$ethnie <- recode_factor(base_tp2$ethnie,  
                                `1` = "Wolof",  
                                `2` = "Polaar/Toucouleur",  
                                `3` = "Sérère",  
                                `4` = "Maindika/Bambara",  
                                `5` = "Soninké",  
                                `6` = "Diola",  
                                `7` = "Manjack",  
                                `8` = "Bainouk",  
                                `9` = "Maures",  
                                `10` = "Balante",  
                                `99` = "Autre")
```

- *Occupation*

```
base_tp2$occupation <- recode_factor( base_tp2$occupation,  
                                       `1` = "Agriculture,Elevage,Sylviculture,pêche",  
                                       `2` = "Activités extractives",  
                                       `3` = "Activité de fabrication(Artisanat)",  
                                       `4` = "Activité de transformation",  
                                       `5` = "Production et distribution d'électricité et de gaz",  
                                       `6` = "Production et distribution d'eau , d'assainissement,
```

- *Formation*

```
base_tp2$formation <- recode_factor(base_tp2$formation,  
                                     `1` = "Non scolarisé",  
                                     `2` = "Elémentaire",  
                                     `3` = "Moyen",  
                                     `4` = "Secondaire",  
                                     `5` = "Master",  
                                     `6` = "Doctorat",
```

```
`99` = "Ne sait pas")
```

- *Niveau d'alphabétisation*

```
base_tp2$niveau_alphabs <- recode_factor(base_tp2$niveau_alphabs,
                                         `0` = "Sans niveau",
                                         `1` = "Sait lire dans une langue",
                                         `2` = "Sait lire et écrire dans une langue")
```

- *Types de variétés* On utilise la fonction `mutate()` de `dplyr` pour appliquer la fonction `labelled()` à chaque variable spécifiée dans `variables`. Les étiquettes *Non* et *Oui* sont attribuées aux valeurs 0 et 1 respectivement.

```
#selection des variables commençant types_varietes_
types = base_tp2 %>%
  dplyr::select(gtsummary::starts_with("types_varietes_")) %>% names()

base_tp2 <- base_tp2 %>%
  dplyr::mutate(across(types,
                        ~ recode_factor(., `0`="Non" , `1`="Oui" )))
```

_critères des variétés

```
#selection des variables commençant critere_var_
criteres = base_tp2 %>%
  dplyr::select(gtsummary::starts_with("criteres_var_")) %>% names()

base_tp2 <- base_tp2 %>%
  dplyr::mutate(across(criteres,
                        ~ recode_factor(., `0`="Non" , `1`="Oui" )))
```

Recodage

2. Statistiques descriptives

- Répartition de la variable sexe

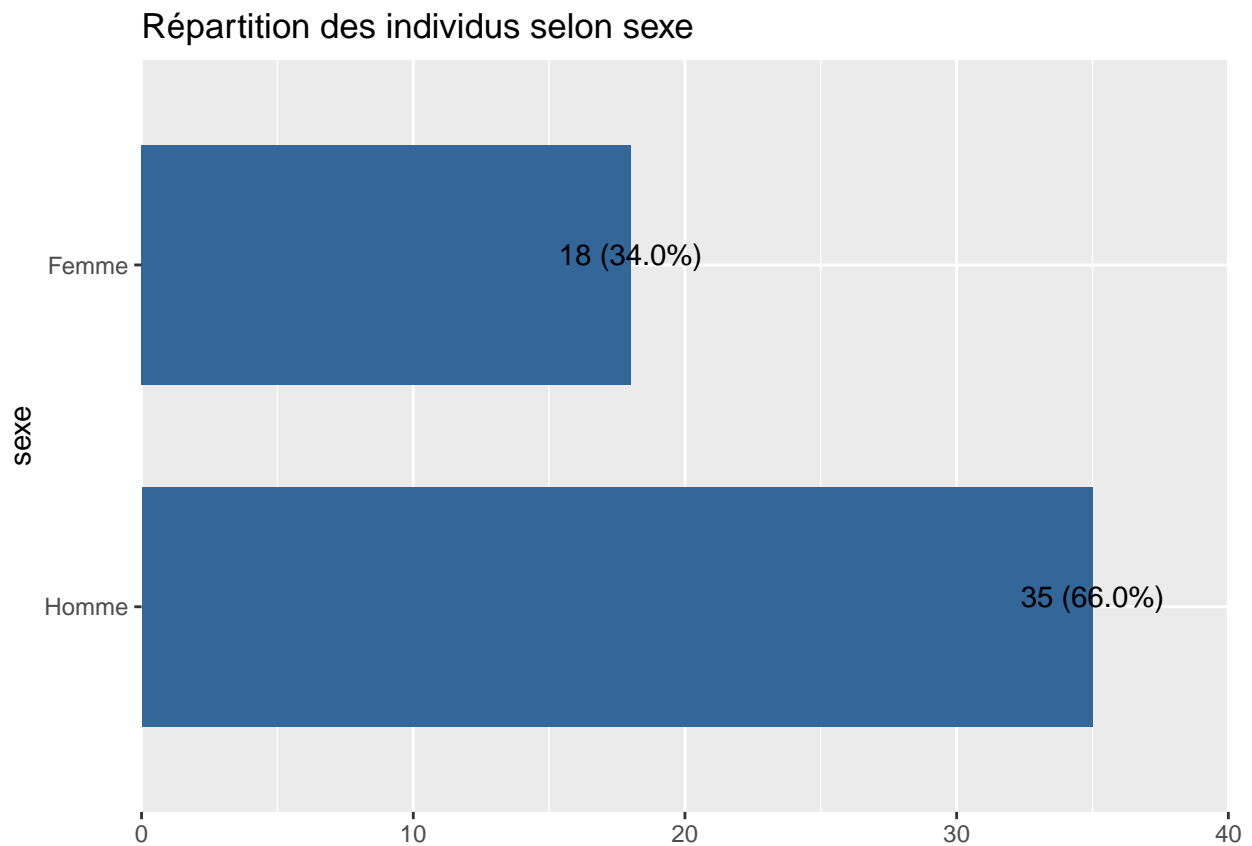
```
#tableau statistique
base_tp2 %>%
  # create a gtsummary table
  tbl_summary(include = "sexe") %>%
  # convert from gtsummary object to gt object
  as_gt() %>%
  # modify with gt functions
  gt::tbl_header("Table 1: répartition du sexe") %>%
  gt::tbl_options(
    table.font.size = "small",
    data_row.padding = gt::px(1))
```

Table 1: répartition du sexe

Characteristic	N = 53 ¹
sexe	
Homme	35 (66%)
Femme	18 (34%)

¹n (%)

```
#diagramme en barre
base_tp2 %>% plot_frq("sexe", coord.flip = T,
  title=paste("Répartition des individus selon", "sexe"), show.na = T)
```

Le tableau ainsi que le graphique présentent la répartition des observations selon le sexe dans un échantillon de taille $N=53$. On constate que les hommes représentent 66% (35 individus) de l'échantillon, tandis que les femmes représentent 34% (18 individus).

- *Situation matrimoniale*

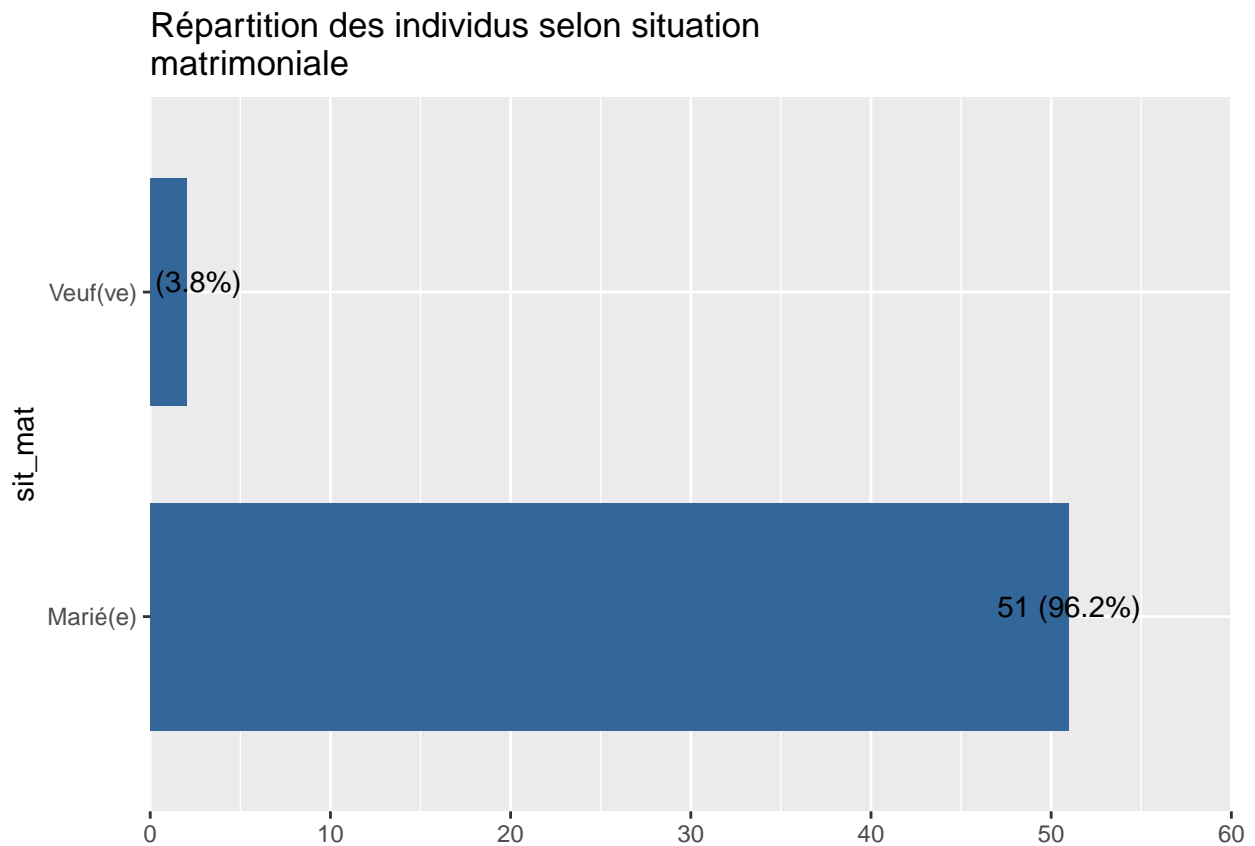
```
#tableau statistique
base_tp2 %>%
  # create a gtsummary table
  tbl_summary(include = "sit_mat") %>%
  # convert from gtsummary object to gt object
  as_gt() %>%
  # modify with gt functions
  gt::tab_header("Table 2: répartition du situation matrimoniale") %>%
  gt::tab_options(
    table.font.size = "small",
    data_row.padding = gt::px(1))
```

Table 2: répartition du situation matrimoniale

Characteristic	N = 53 ¹
sit_mat	
Marié(e)	51 (96%)
Veuf(ve)	2 (3.8%)

¹n (%)

```
#diagramme en barre
base_tp2 %>% plot_frq("sit_mat", coord.flip = T,
                      title=paste("Répartition des individus selon", "situation matrimoniale"))
```



Le tableau et le graphique révèlent que la majorité des individus (96%, soit 51 personnes) sont marié(e)s, tandis qu'une minorité (3.8%, soit 2 personnes) sont veuf(ve)s.

- *Formation*

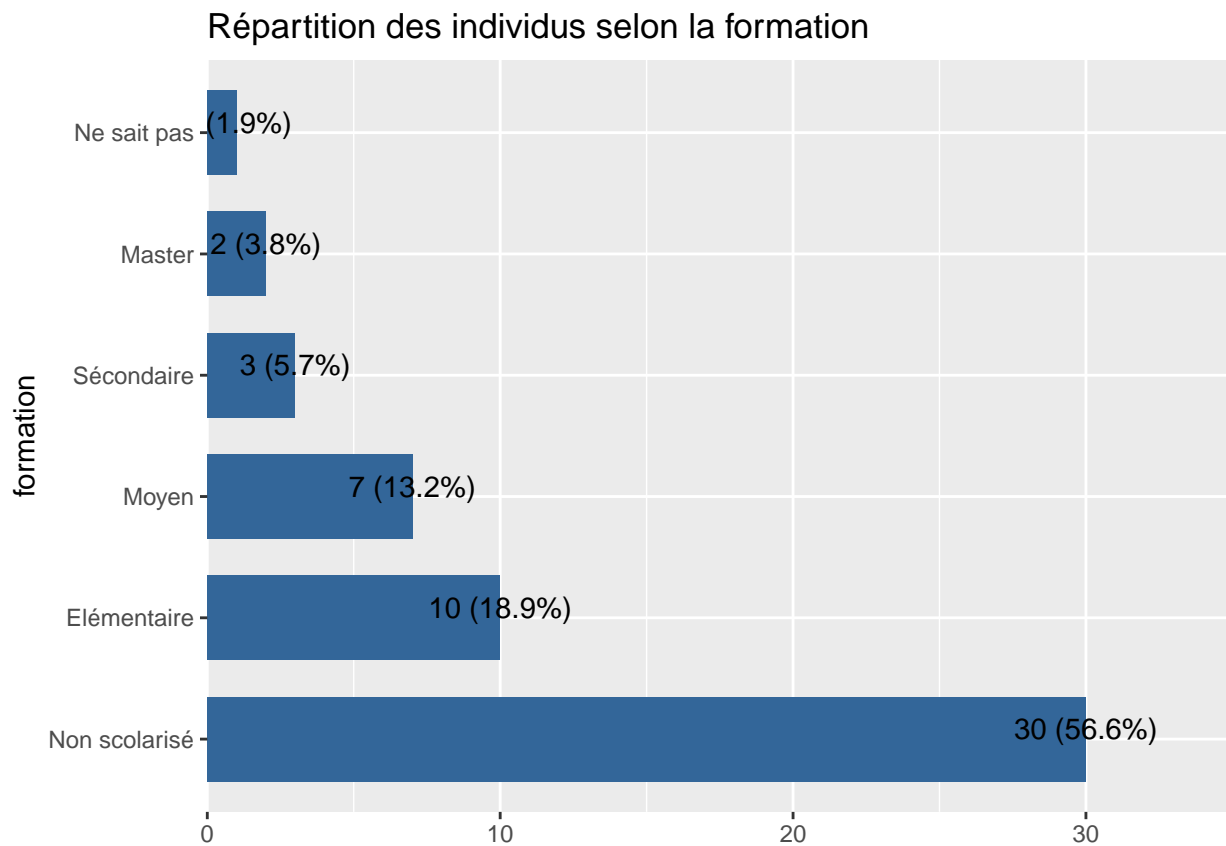
```
base_tp2 %>%
  # create a gtsummary table
  tbl_summary(include = "formation") %>%
  # convert from gtsummary object to gt object
  as_gt() %>%
  # modify with gt functions
  gt::tab_header("Table 3: répartition formation") %>%
  gt::tab_options(
    table.font.size = "small",
    data_row.padding = gt::px(1))
```

Table 3: répartition formation

Characteristic	N = 53 ¹
formation	
Non scolarisé	30 (57%)
Elémentaire	10 (19%)
Moyen	7 (13%)
Sécondaire	3 (5.7%)
Master	2 (3.8%)
Ne sait pas	1 (1.9%)

¹n (%)

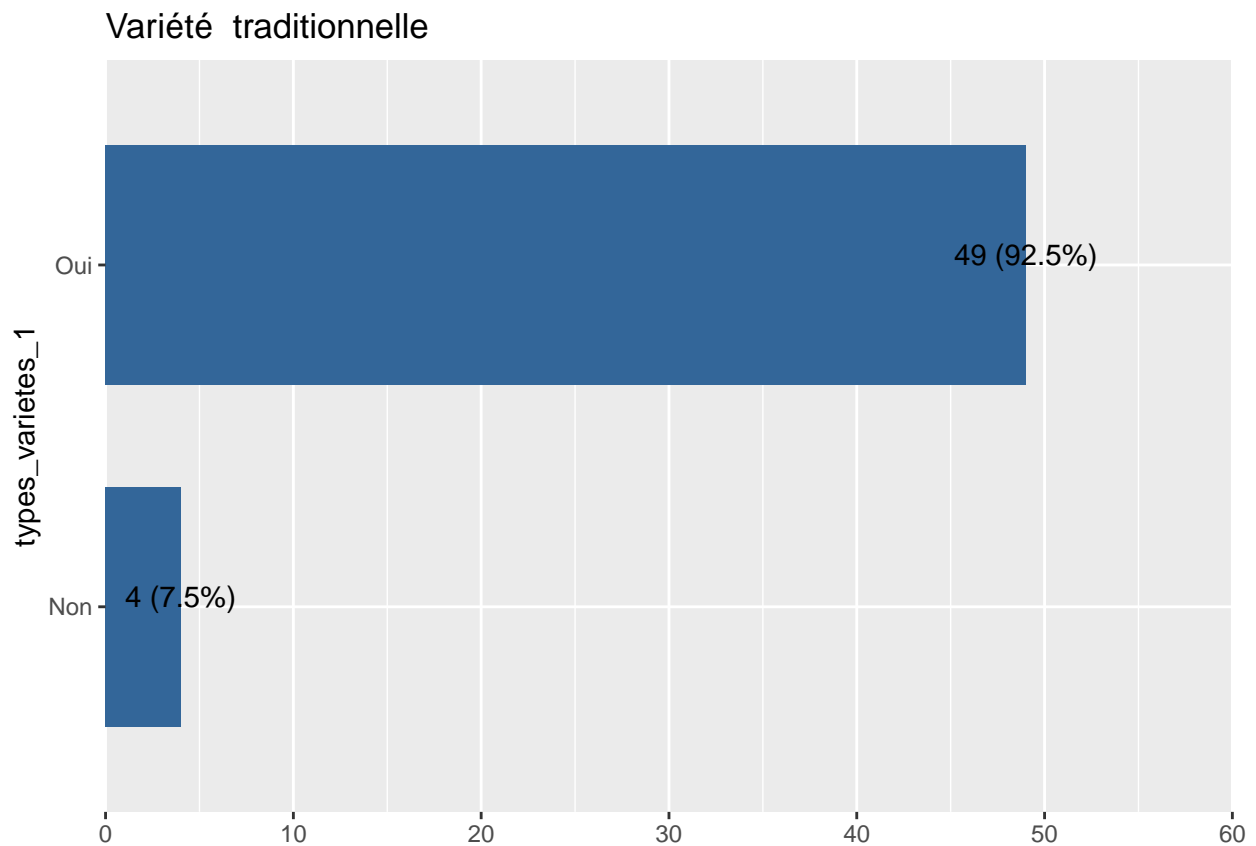
```
#diagramme en barre
base_tp2 %>%
  plot_frq("formation", coord.flip = T,
    title="Répartition des individus selon la formation", show.na = T)
```



On constate que la majorité des individus ont un niveau de formation “Non scolarisé” (57%, soit 30 personnes), suivis par “Elémentaire” (19%, soit 10 personnes), “Moyen” (13%, soit 7 personnes), “Secondaire” (5.7%, soit 3 personnes), et “Master” (3.8%, soit 2 personnes). Seule une petite fraction de l’échantillon (1.9%, soit 1 personne) ne sait pas quel est son niveau de formation.

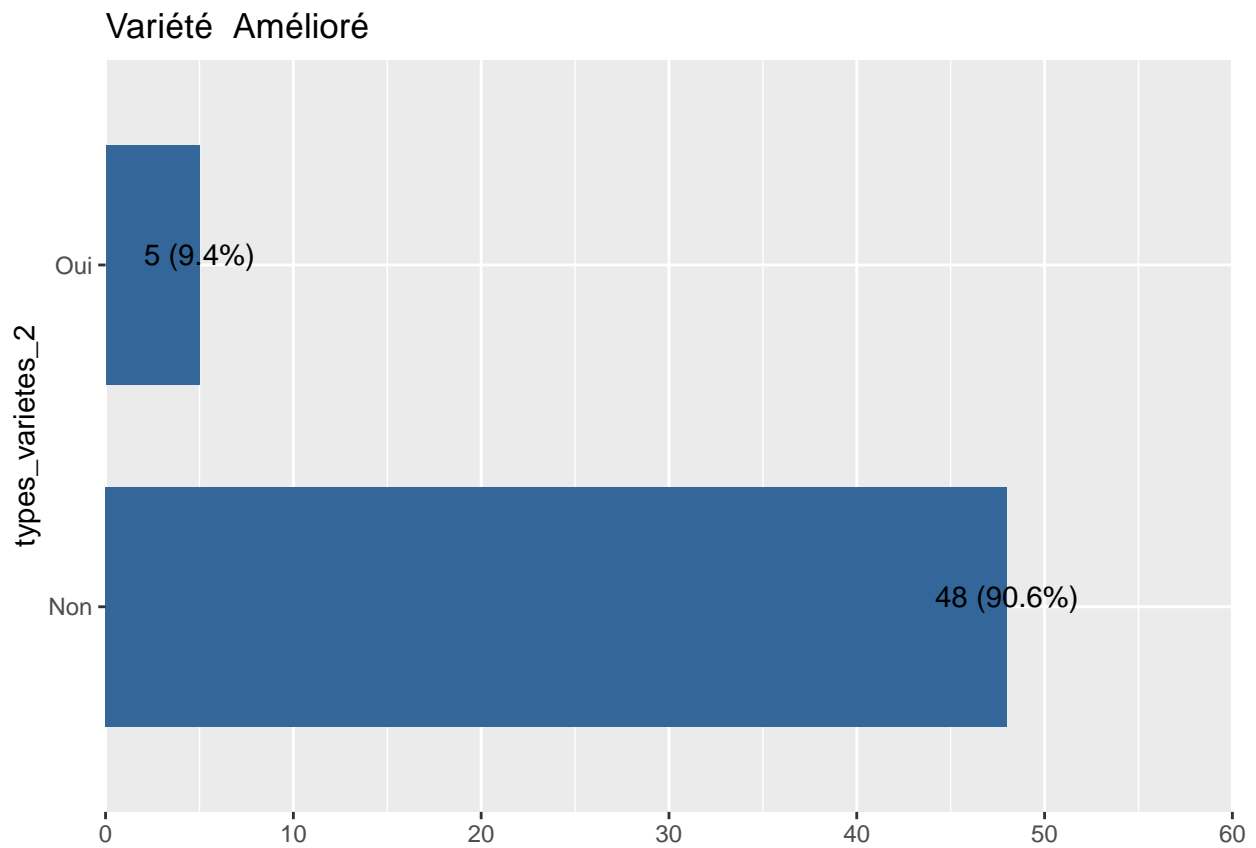
- *Types de variétés de sésames utilisés pour la production*

```
base_tp2 %>% plot_frq("types_varietes_1", coord.flip = T,
                     title="Variété traditionnelle", show.na = T)
```



Une analyse du graphique montre que 92,5% des individus de notre échantillon utilise les sesames traditionnelles pour la production contre 7,5%.

```
base_tp2 %>% plot_frq("types_varietes_2", coord.flip = T,  
                      title="Variété Amélioré", show.na = T)
```



Une autre analyse montre que seulement 9,4% des individus utilise la variété amélioré du sésame contre 90,6%.