|  |
| --- |
| **REPUBLIQUE DU SENEGAL** |
|  |
| **\*\*\*\*\*\*\*\*\*\*** |
| ***Un Peuple - Un But - Une Foi*** |
| **\*\*\*\*\*\*\*\*\*\*** |
| **Agence nationale de la Statistique et de la démographie** |
|  |
| **\*\*\*\*\*\*\*\*\*\*** |
| **Ecole nationale de la Statistique et de l'Analyse économique Pierre Ndiaye** |
|  |

##### Projet statistique sur R : Evaluation

|  |  |
| --- | --- |
| **Rédigé par** | **Sous la supervision de** |
| KAFANDO G. Judicaël Oscar | M. Aboubacar HEMA |
| *Élève Ingénieur Statisticien Économiste* | *Data-scientist* |
| **Année académique 2024-2025** | |

###### Sommaire

[I. Analayse et compréhension de la base 7](#_Toc198042568)

[1. Structure de la base 7](#_Toc198042569)

[2. Vérification et suppression des colonnes ou lignes vides 7](#_Toc198042570)

[3. Vérifications des doublons 7](#_Toc198042571)

[4. Vérification des identifiants (cohérence relationnelle) 7](#_Toc198042572)

[5. Vérification de la cohérence inter-variable 8](#_Toc198042573)

[6. Vérification des valeurs manquantes et aberrantes 8](#_Toc198042574)

[II. Caractéristiques socio-démographiques 8](#_Toc198042575)

[7. Pyramide des ages des indivdus 8](#_Toc198042576)

[a. Préparation des données 8](#_Toc198042577)

[b. Representation de la pyramide 8](#_Toc198042578)

[c. Caractéristique du chef de ménage 9](#_Toc198042579)

[d. Calculez le nombre d’individu par menage 10](#_Toc198042580)

[e. Faites les statistiques descriptives sur cette variable 10](#_Toc198042581)

[f. Faites des statistiques descriptives sur la variable renseigant le nombre de pieces du menage 11](#_Toc198042582)

[g. Calculez l’indice d’affluence ou crowding index 11](#_Toc198042583)

[h. Quelle est la proportion de menage dont le crowding index est: <1, 1<= crowding\_index<2, 2<= crowding\_index<3 et >=3 12](#_Toc198042584)

[i. Comparaison entre réfugiés et communautés d’accueil 13](#_Toc198042585)

[III. Analyse de la securite alimentaire des deplaces internes 13](#_Toc198042586)

[8. Score de consommation alimentaire (SCA): 10 points 13](#_Toc198042587)

[j. Analyse descriptives des variables qui compose le SCA 14](#_Toc198042588)

[k. Calculer le score de consommation alimentaire 1](#_Toc198042589)

[l. Faites un tableau illustrant le poids attribue a chaque groupe alimentaire pour le calcul du SCA (la somme totale des poids doit etre egale a 16) 1](#_Toc198042590)

[m. Categorisation du SCA selon les seuil (21 et 35 / 28 et 42) 1](#_Toc198042591)

[n. Faites une répresentation spatiale (région et département) du SCA et de ses différentes catégorisations. 2](#_Toc198042592)

[o. Moyenne du SCA par région : 2](#_Toc198042593)

[p. Vérification des noms des régions 2](#_Toc198042594)

###### Liste des figures

[**Figure :** Pyramide des âges par sexe 9](#_Toc198042595)

[**Figure :** Distribution de l’indice d’affluence (Crowding Index) 12](#_Toc198042596)

[**Figure :** Proportion des ménages par catégorie d’indice d’affluence 13](#_Toc198042597)

[**Figure :** Distribution de l’indice d’affluence par groupe de population 13](#_Toc198042598)

[**Figure :** Carte du Score de Consommation Alimentaire (SCA) par région 3](#_Toc198042599)

[**Figure :** Carte des catégories du SCA par région 3](#_Toc198042600)

###### Liste des tableaux

[Tableau : Caractéristiques des individus selon le groupe 9](#_Toc198042601)

[Tableau : Statistiques descriptives de la taille du ménage 10](#_Toc198042602)

[Tableau : Statistiques descriptives du nombre de pièce 11](#_Toc198042603)

[Tableau : Effectifs et pourcentages des groupes alimentaires selon le nombre de jours (0–7) 15](#_Toc198042604)

[Tableau : Poids attribués aux groupes alimentaires pour le calcul du Food Consumption Score (FCS) 1](#_Toc198042605)

[Tableau : Aperçu des 10 premières observations du score de consommation alimentaire (SCA) 1](#_Toc198042606)

###### Introduction

Dans le cadre du cours de Projet statistique sous R, cette étude applique les compétences acquises en traitement, analyse et visualisation de données pour explorer la situation alimentaire et nutritionnelle des populations réfugiées et des communautés hôtes. À travers l’exploitation d’une base de données issue d’une enquête terrain, ce travail vise à décrire les caractéristiques socio-démographiques des ménages, analyser la diversité alimentaire et identifier les vulnérabilités spécifiques selon les groupes. L’ensemble des analyses, réalisées sous R, permet d’illustrer concrètement l’usage d’outils statistiques dans un contexte humanitaire, et de fournir des résultats appuyés sur des méthodes rigoureuses et reproductibles.

# Analayse et compréhension de la base

## Structure de la base

Le but est d’avoir une vue globale de la base.

La base individu contient 144 variables sur 22092 individus.  
Parmi les variables, on trouve : 9 variable(s) du type character, 2 variable(s) du type Date, 60 variable(s) du type haven\_labelled, 73 variable(s) du type numeric.

La base principale contient 1312 variables sur 3058 individus.  
Parmi les variables, on trouve : 53 variable(s) du type character, 2 variable(s) du type Date, 362 variable(s) du type haven\_labelled, 895 variable(s) du type numeric.

## Vérification et suppression des colonnes ou lignes vides

Les **colonnes et lignes vides** n’ont aucune information pour nous. Donne nous les supprimons pour alleger la base et le traitement.

Les bases **principale et individus** contiennent respectivement 68 et 0 colonnes vides et 0 et 0lignes vides.

## Vérifications des doublons

Les doublons dans une base, compte pour des informations répétées et qui biaisent les résultats. Nous examinons donc les deux bases et en cas d’existence, nous les supprimons pour ne garder qu’un seul.

La base principale a 0 doublons et la base individu en a 0.

## Vérification des identifiants (cohérence relationnelle)

Les vérifications effectuées sur les identifiants ont permis de confirmer l’intégrité relationnelle entre les deux bases de données. Aucun doublon n’a été détecté au niveau des identifiants des ménages, ce qui garantit leur unicité. De plus, l’ensemble des individus recensés dans la base individuelle sont correctement rattachés à un ménage valide dans la base principale. Enfin, la combinaison de l’identifiant du ménage et de la position dans le ménage **(rosterposition)** est unique, ce qui atteste de l’absence de doublons au sein des ménages. Ces résultats valident la structure hiérarchique des données et confirment la fiabilité des liens entre les deux niveaux d’observation.

## Vérification de la cohérence inter-variable

L’**objectif** est de s’assurer que les réponses d’une variable sont cohérentes avec d’autres réponses du même individu ou ménage. Par exemple :

* *Une personne qui dit ne pas avoir fui ne peut pas déclarer une cause de fuite*
* *Une personne ayant déclaré être née dans le pays d’accueil ne devrait pas avoir une date d’arrivée*
* *Une femme de 5 ans ayant eu un enfant : incohérent*

Les résultats des trois contrôles ciblés montre que chez les personnes ayant déclaré ne pas avoir fui leur domicile et ceux n’avaient pas indiqué de raison de fuite ; aucune contradiction n’a été observée. Pour les individus ayant déclaré avoir toujours vécu dans le pays d’accueil tout en renseignant une date d’arrivée, aucun cas de ce type n’a été trouvé. Enfin, la présence d’enfants de moins de 10 ans ayant déclaré avoir donné naissance à un enfant, aucune anomalie n’a été détectée.

## Vérification des valeurs manquantes et aberrantes

Objectif : s’assurer que les variables importantes sont renseignées et qu’il n’y a pas d’aberrations statistiques ou logiques.

La vérification des âges a permis de s’assurer qu’aucune valeur aberrante ne figure dans la variable age\_years. Aucun individu n’a été enregistré avec un âge négatif ou supérieur à 120 ans. Aussi, aucune incohérence n’est observée,au niveau de la variable sexe. Enfin, l’analyse de la distribution des réponses permet de confirmer la diversité attendue des positions dans le ménage (chef, conjoint(e), enfant, etc.). Aucun code inattendu n’a été détecté, et le nombre de chef de ménage est égale à la taille de la base ménage.

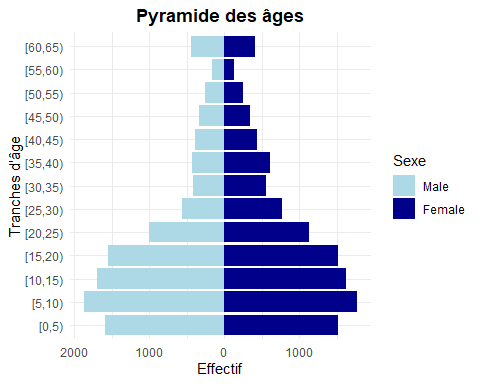
# Caractéristiques socio-démographiques

## Pyramide des ages des indivdus

### Préparation des données

Dans cette partie, nous préparons les données pour une meilleur représentation

### Representation de la pyramide



**Figure** **1:** Pyramide des âges par sexe

La pyramide des âges issue de l’enquête sur les déplacements forcés au Sud-Soudan en 2023 met en évidence une structure démographique fortement juvénile, caractérisée par une concentration marquée dans les tranches d’âge de 0 à 20 ans. Cette surreprésentation des enfants et adolescents reflète les dynamiques de natalité élevées typiques des contextes de crise humanitaire. La distribution par sexe reste relativement équilibrée dans les tranches jeunes, mais l’on observe une prédominance féminine progressive chez les adultes, probablement liée à des schémas de mobilité, de mortalité différentielle ou de composition des foyers déplacés. La faible présence des personnes âgées et quelques pics irréguliers dans les classes supérieures suggèrent à la fois une espérance de vie limitée et des imprécisions dans la déclaration de l’âge.

### Caractéristique du chef de ménage

Tableau 1: Caractéristiques des individus selon le groupe

| Caractéristiques du chef de menage | **Refugees**  N = 2,0681 | **Host community North**  N = 9901 | **p-value**2 |
| --- | --- | --- | --- |
| **age\_years** | 41.6 (17.0,61.0)) | 41.6 (19.0,61.0)) | >0.9 |
| **hh\_02** |  |  | <0.001 |
| Male | 54% | 63% |  |
| Female | 46% | 37% |  |
| **hh\_08** |  |  |  |
| monogamous/married | 63% | 72% |  |
| polygamous/married | 14% | 19% |  |
| non-formal union | 0.6% | 0.3% |  |
| separated | 5.4% | 1.5% |  |
| divorced | 3.3% | 0.9% |  |
| widow or widower | 7.4% | 5.2% |  |
| never married | 6.9% | 1.1% |  |
| 1Mean (Min,Max)); % | | | |
| 2Welch Two Sample t-test; Pearson's Chi-squared test | | | |

Les chefs de ménage réfugiés et ceux de la communauté hôte du Nord ont un âge médian similaire (environ 41,6 ans). Cependant, on observe des différences significatives selon le sexe : les femmes représentent 46 % chez les réfugiés contre 37 % dans la communauté hôte (p < 0,001). Concernant le statut matrimonial, les réfugiés sont moins souvent mariés de façon monogame (63 % vs 72 %) ou polygame (14 % vs 19 %). En revanche, ils sont plus souvent séparés, divorcés, veufs ou jamais mariés que les membres de la communauté hôte. Ces différences sont statistiquement significatives.

##Crowding Index ou l’indice d’affluence

L’indice d’affluence est le nombre de membres du menage divise par le nombre de pièces (à l’exclusion de la cuisine et des couloirs).

### Calculez le nombre d’individu par menage

Pour cette partie, nous utilisons la base individu et nous sommons simplement les indivdus du meme ménage pour avoir la taille du ménage. Une fois cela fait, nous ajoutons cette variable à la base principale.

### Faites les statistiques descriptives sur cette variable

Tableau 2: Statistiques descriptives de la taille du ménage

| Statistique | Valeur |
| --- | --- |
| Moyenne | 7.2 |
| Médiane | 7.0 |
| Mode | 7.0 |
| Écart-type | 3.2 |
| Minimum | 1.0 |
| Maximum | 17.0 |

La taille moyenne des ménages est de 7,2 personnes, avec une médiane et un mode de 7, ce qui indique une distribution centrée autour de cette valeur. La taille des ménages varie de 1 à 17 personnes, avec un écart-type de 3,2, montrant une certaine variabilité entre les ménages.

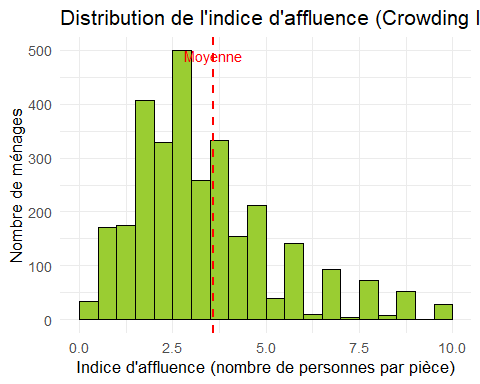
### Faites des statistiques descriptives sur la variable renseigant le nombre de pieces du menage

Tableau 3: Statistiques descriptives du nombre de pièce

| Statistique | Valeur |
| --- | --- |
| Moyenne | 2.4 |
| Médiane | 2.0 |
| Mode | 2.0 |
| Écart-type | 1.3 |
| Minimum | 1.0 |
| Maximum | 21.0 |

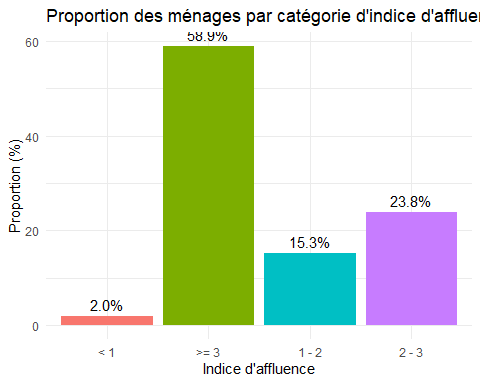
En moyenne, les ménages disposent de 2,4 pièces, avec une médiane et un mode de 2, ce qui suggère que la plupart des logements ont deux pièces. Le nombre de pièces varie de 1 à 21, avec un écart-type de 1,3, indiquant une dispersion modérée autour de la moyenne.

### Calculez l’indice d’affluence ou crowding index



**Figure** **2:** Distribution de l’indice d’affluence (Crowding Index)

### Quelle est la proportion de menage dont le crowding index est: <1, 1<= crowding\_index<2, 2<= crowding\_index<3 et >=3

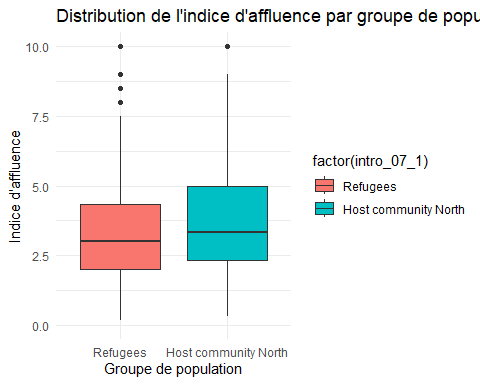


**Figure** **3:** Proportion des ménages par catégorie d’indice d’affluence

La majorité des ménages (58,9 %) appartiennent à la catégorie d’affluence élevée (≥ 3), suivie de 23,8 % dans la catégorie 2–3 et de 15,3 % dans la catégorie 1–2. Seuls 2 % des ménages ont une affluence inférieure à 1, indiquant que très peu de ménages sont faiblement peuplés.

### Comparaison entre réfugiés et communautés d’accueil

#### Boite à moustache



**Figure** **4:** Distribution de l’indice d’affluence par groupe de population

Le graphique montre qu’en moyenne, **les membres de la communauté hôte du Nord présentent un indice légèrement plus élevé que les réfugiés**. La médiane est aussi un peu plus haute chez la communauté hôte. La variabilité des scores est plus grande dans ce groupe, avec quelques valeurs extrêmes dans les deux groupes. Cela suggère que les conditions d’affluence sont globalement plus favorables chez la communauté hôte, mais restent hétérogènes dans les deux cas.

# Analyse de la securite alimentaire des deplaces internes

## Score de consommation alimentaire (SCA): 10 points

### Analyse descriptives des variables qui compose le SCA

Tableau 4: Effectifs et pourcentages des groupes alimentaires selon le nombre de jours (0–7)

| **jours** | **food\_div1** | **food\_div2** | **food\_div3** | **food\_div4** | **food\_div5** | **food\_div6** | **food\_div7** | **food\_div8** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **0** | 677 (22.1%) | 1900 (62.1%) | 2479 (81.1%) | 2179 (71.3%) | 2099 (68.6%) | 2775 (90.7%) | 2144 (70.1%) | 2280 (74.6%) |
| **1** | 450 (14.7%) | 317 (10.4%) | 228 (7.5%) | 397 (13%) | 244 (8%) | 56 (1.8%) | 152 (5%) | 194 (6.3%) |
| **2** | 253 (8.3%) | 276 (9%) | 113 (3.7%) | 231 (7.6%) | 243 (7.9%) | 56 (1.8%) | 198 (6.5%) | 161 (5.3%) |
| **3** | 365 (11.9%) | 228 (7.5%) | 79 (2.6%) | 124 (4.1%) | 175 (5.7%) | 47 (1.5%) | 199 (6.5%) | 129 (4.2%) |
| **4** | 230 (7.5%) | 89 (2.9%) | 27 (0.9%) | 34 (1.1%) | 91 (3%) | 25 (0.8%) | 109 (3.6%) | 73 (2.4%) |
| **5** | 297 (9.7%) | 86 (2.8%) | 38 (1.2%) | 31 (1%) | 64 (2.1%) | 38 (1.2%) | 70 (2.3%) | 66 (2.2%) |
| **6** | 92 (3%) | 43 (1.4%) | 15 (0.5%) | 12 (0.4%) | 15 (0.5%) | 3 (0.1%) | 41 (1.3%) | 33 (1.1%) |
| **7** | 690 (22.6%) | 113 (3.7%) | 72 (2.4%) | 40 (1.3%) | 112 (3.7%) | 42 (1.4%) | 134 (4.4%) | 106 (3.5%) |
|  | 4 (0.1%) | 6 (0.2%) | 7 (0.2%) | 10 (0.3%) | 15 (0.5%) | 16 (0.5%) | 11 (0.4%) | 16 (0.5%) |

Les résultats montrent que la consommation alimentaire varie largement selon les groupes. Les céréales et tubercules (food\_div1) sont les plus fréquemment consommés, avec 22,1 % des ménages n’en consommant pas du tout, et 22,6 % les consommant quotidiennement. En revanche, les légumineuses/noix (food\_div2), les légumes (food\_div5) et les fruits (food\_div6) sont moins présents, avec des proportions élevées de ménages ne les consommant pas du tout (respectivement 62,1 %, 81,1 %, et 90,7 %). La consommation de viande/poisson/œufs (food\_div4) et de produits laitiers (food\_div3) reste également faible, bien que plus fréquente que celle des légumes et fruits. Enfin, les matières grasses (food\_div7) et le sucre (food\_div8) sont peu consommés au quotidien, avec de faibles proportions de ménages les intégrant régulièrement dans leur alimentation.

### Calculer le score de consommation alimentaire

🎯 Objectif Le SCA est obtenu en multipliant la fréquence de consommation de chaque groupe alimentaire par un poids nutritionnel, puis en faisant la somme pondérée

### Faites un tableau illustrant le poids attribue a chaque groupe alimentaire pour le calcul du SCA (la somme totale des poids doit etre egale a 16)

Tableau 5: Poids attribués aux groupes alimentaires pour le calcul du Food Consumption Score (FCS)

| **Variable** | **Groupe alimentaire** | **Poids attribué** |
| --- | --- | --- |
| food\_div1 | Céréales & tubercules (jours) | 2.0 |
| food\_div2 | Légumineuses/noix (jours) | 3.0 |
| food\_div5 | Légumes (jours) | 1.0 |
| food\_div6 | Fruits (jours) | 1.0 |
| food\_div4 | Viande/poisson/œufs (jours) | 4.0 |
| food\_div3 | Produits laitiers (jours) | 4.0 |
| food\_div8 | Sucre (jours) | 0.5 |
| food\_div7 | Matières grasses (jours) | 0.5 |
|  | Total | 16.0 |

### Categorisation du SCA selon les seuil (21 et 35 / 28 et 42)

Tableau 6: Aperçu des 10 premières observations du score de consommation alimentaire (SCA)

| **Household ID** | **SCA** | **SCA\_cat1** | **SCA\_cat2** |
| --- | --- | --- | --- |
| 1 | 0.0 | Faible | Faible |
| 2 | 10.0 | Faible | Faible |
| 3 | 9.0 | Faible | Faible |
| 4 | 6.0 | Faible | Faible |
| 5 | 37.0 | Acceptable | Limite |
| 6 | 2.5 | Faible | Faible |
| 7 | 50.5 | Acceptable | Acceptable |
| 8 | 2.0 | Faible | Faible |
| 9 | 13.0 | Faible | Faible |
| 10 | 9.5 | Faible | Faible |

Le tableau ci-dessus, resume les différents score SCA mais pour quelques individus de la base.

### Faites une répresentation spatiale (région et département) du SCA et de ses différentes catégorisations.

🎯 Objectif : Visualiser la moyenne du SCA ou la part des ménages par catégorie dans chaque région ou département.

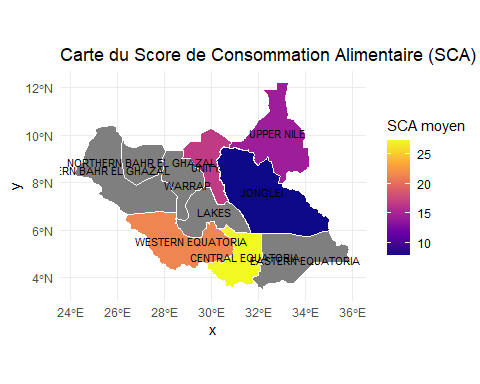
### Moyenne du SCA par région :

### Vérification des noms des régions

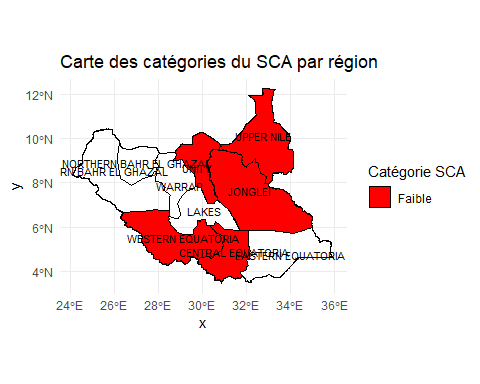
[1] NA

Score moyen (SCA) de chaque région

| **admin1** | **SCA\_moyen** | **SCA\_moyen\_cat2** |
| --- | --- | --- |
| **CENTRAL EQUATORIA** | 27.27434 | Faible |
| **JONGLEI** | 8.00000 | Faible |
| **UNITY** | 16.76490 | Faible |
| **UPPER NILE** | 14.50959 | Faible |
| **WESTERN EQUATORIA** | 21.41404 | Faible |
|  | 17.98370 | Faible |



**Figure** **5:** Carte du Score de Consommation Alimentaire (SCA) par région



**Figure** **6:** Carte des catégories du SCA par région

Cette carte révèle que parmi les régions couvertes par l’enquête, aucune n’affiche un score moyen ou élevé. L’ensemble des régions présentent uniquement des scores faibles, traduisant une situation globalement préoccupante.