Ch10 - WS 1.1 - 1.3

Reid Minto

April 22, 2020

1. Producing two more tables: ws1.2 and ws1.3 as well
2. Integrate in the code suppression and parenthesizing rules:
3. Where the unweighted count of the denominator is less than 25 cases the value should be suppressed and replaced with an asterisk \*
4. Where the unweighted count is between 25-49 the values should be parenthesized ().
5. Compile the three tables’ scripts into 1 markdown that can produce an SFR chapter (albeit for only 3 tables) in word document
6. It will be paramount here that the table formatting in the original SFR reports are maintained.

library(tidyverse); library(haven); library(lubridate);  
library(knitr); library(kableExtra);  
library(Hmisc); library(survey)  
library(expss); library(here); library(gmodels)  
library(flextable); library(formattable)  
library(prettydoc); library(officer)  
  
library(tableone); library(sjPlot); library(srvyr)

# Import data

here function navigates you to top level Rproj directory

# haven library called above has read\_sav function for reading in .sav / .sps etc  
# files from other languages  
here()

## [1] "C:/Users/Owner/Documents/UNICEF-MICS"

hh <- read\_sav(here("Data/Tunisia\_hh.sav"))  
# hh <- as\_factor(hh, only\_labelled = TRUE)

Filter only cases where hh interview = 1, create a variable hhweightHH48 by multiplying HH48 and hhweight, number of hh members by sample weight

Note: this is just first pass rough draft / run through - much of this looks like it can be consolidated into a single chunk / a single function execute per section from what I’ve seen so far, but broken up in first pass

# 10 - WS.1.1

hh <- hh %>% filter(HH46 == 1) %>%   
 mutate(hhweightHH48 = HH48 \* hhweight) %>% mutate(nhhmem = 1 \* hhweightHH48)  
  
var\_lab(hh$nhhmem) = "Number of household members"; val\_lab(hh$nhhmem) = make\_labels("  
 1  
 ")  
  
  
  
var\_lab(hh$WS1) = "Source principale d'eau de boisson"  
val\_lab(hh$WS1) = make\_labels("  
 11 DANS LE LOGEMENT  
 12 DANS LA CONCESSION/JARDIN/ PARCELLE  
 13 CHEZ LE VOISIN  
 14 ROBIENT PUBLIC/BORNE FONTAINE  
 21 PUITS A POMPE/FORAGE  
 31 PROTEGE  
 32 PAS PROTEGE  
 41 SOURCE PROTEGEE  
 42 SOURCE NON PROTEGEE  
 51 EAU DE PLUIE  
 61 CAMION CITERNE  
 71 CHARRETTE AVEC PETITE CITERNE  
 81 EAU DE SURFACE (RIVIERE, BARRAGE, LAC, MARE, COURANT, CANAL, SYSTEME D’IRRIGATION)  
 91 EAU EN BOUTEILLE  
 92 EAU EN SACHET  
 96 AUTRE  
 99 NON REPONSE  
 ")  
############ Begin 10 - WS.sps ############################  
  
hh <- hh %>%   
 mutate(drinkingWater = ifelse(WS1 %in% c(11:14, 21, 31, 41, 51, 61, 71, 72, 91, 92), 1, 2))  
  
var\_lab(hh$drinkingWater) = "Main source of drinking water"  
val\_lab(hh$drinkingWater) = make\_labels("  
 1 Improved sources  
 2 Unimproved sources  
")  
  
  
  
hh <- hh %>%   
 mutate(toiletType = case\_when(  
 WS11 %in% c(11:13, 18, 21, 22, 31) ~ 1,  
 WS11 == 95 ~ 3,  
 TRUE ~ 2  
 ))  
  
  
var\_lab(hh$toiletType) = "Type of sanitation facility"  
val\_lab(hh$toiletType) = make\_labels("  
 1 Improved  
 2 Unimproved  
 3 Open defecation (no facility, bush, field)  
 ")  
  
  
hh <- hh %>% mutate(flush = NA) %>% mutate(flush = case\_when(  
 WS11 %in% c(1:14, 18) ~ 1  
))  
  
var\_lab(hh$flush) = ""; val\_lab(hh$flush) = make\_labels("   
 1 Flush/Pour flush to:  
 ")  
  
  
  
hh <- hh %>% mutate(sharedToilet = case\_when(  
 WS17 %in% c(1:5) ~ 1,  
 WS17 %in% 97:99 ~ 9,  
 NA ~ 0,  
 TRUE ~ 2  
))  
  
hh <- hh %>% mutate(sharedToilet = ifelse(WS16 == 2, 3, sharedToilet))  
  
var\_lab(hh$sharedToilet) = " "  
val\_lab(hh$sharedToilet) = make\_labels("  
 0 Not shared  
 1 5 households or less  
 2 More than 5 households  
 3 Public facility  
 9 DK/Missing  
 ")  
  
  
############ End 10 - WS.sps ############################  
  
  
hh <- hh %>% mutate(improvedWater = ifelse(drinkingWater == 1, 100, 0))   
  
  
var\_lab(hh$improvedWater) = "Percentage using improved sources of drinking water [1]"  
var\_lab(hh$WS1) = ""  
  
  
hh <- hh %>% mutate(total = 1, total100 = 100)  
var\_lab(hh$total) = "Total"; var\_lab(hh$total100) = "Total"  
val\_lab(hh$total) = make\_labels("   
 1  
 ")  
val\_lab(hh$total100) = make\_labels("  
 100  
 ")

# 10.1.1 Table

tab.1.1 = xtabs(nhhmem ~ WS1, hh) %>%   
 rbind(xtabs(nhhmem ~ as\_factor(HH6) + WS1, hh)) %>%   
 rbind(xtabs(nhhmem ~ as\_factor(HH7) + WS1, hh)) %>%  
 rbind(xtabs(nhhmem ~ as\_factor(helevel) + WS1, hh)) %>%   
 rbind(xtabs(nhhmem ~ as\_factor(windex5) + WS1, hh))  
  
  
tab.improved = xtabs(nhhmem ~ improvedWater, hh) %>%   
 rbind(xtabs(nhhmem ~ as\_factor(HH6) + improvedWater, hh)) %>%   
 rbind(xtabs(nhhmem ~ as\_factor(HH7) + improvedWater, hh)) %>%  
 rbind(xtabs(nhhmem ~ as\_factor(helevel) + improvedWater, hh)) %>%   
 rbind(xtabs(nhhmem ~ as\_factor(windex5) + improvedWater, hh))  
  
  
tab.1.1 = tab.1.1 %>%  
 cbind(Total = rowSums(tab.1.1)) %>%   
 cbind(tab.improved[, 2])  
  
N = round(tab.1.1[ , 17])  
  
colnames(tab.1.1)[18] = "Percentage using drinking water from improved sources"  
  
  
  
#labels(tabf)[[1]][1] = "Total"  
tab.1.1 = apply(tab.1.1, 2, function(x) (x\*100) / tab.1.1[, 17])  
  
tab.1.1 <- tab.1.1 %>% cbind(N); colnames(tab.1.1)[18:19] = c("Pourcen-tage utilisant de l'eau de boisson de sources améliorées1", "Nombre de membres des ménages ")  
  
rownames(tab.1.1)[1] = "Total"  
  
  
  
  
  
# columns are labeled incorrectly   
# SPSS syntax does not match table output, even though I get correct results, they are nested in   
# incorrect value labels / label hierarchies, can fix this as I go, need to investigate though  
  
ind <- c(1, 2, 3, 4, 5, 6, 8, 10, 11, 12, 14, 15, 7, 9, 13, 16, 17)  
  
tab.1.1[ , 1:17] = tab.1.1[ , ind]  
  
colnames(tab.1.1)[1:17] = colnames(tab.1.1)[ind]

# SFR Table Ch10 WS 1.1

tab.1.1 %>%   
 kable(digits = 1, format.args = list(big.mark = ",", scientific = FALSE)) %>%   
 kable\_styling(c("striped"), font\_size = 11) %>%   
# pack\_rows("Total", 1, 1) %>%   
 pack\_rows(label(hh$HH6), 2, 3) %>%   
 pack\_rows(label(hh$HH7), 4, 10) %>%   
 pack\_rows(label(hh$helevel), 11, 15) %>%   
 pack\_rows(label(hh$windex5), 16, 20) %>%   
 add\_header\_above(c(" " , "Eau courante" = 4, " " = 15)) %>%   
 add\_header\_above(c(" ", "Sources améliorées" = 12, "Sources non améliorées" = 4, " " = 3)) %>%   
 add\_header\_above(c(" ", "Source principale d'eau de boisson" = 19))

Source principale d’eau de boisson

Sources améliorées

Sources non améliorées

Eau courante

DANS LE LOGEMENT

DANS LA CONCESSION/JARDIN/ PARCELLE

CHEZ LE VOISIN

ROBIENT PUBLIC/BORNE FONTAINE

PUITS A POMPE/FORAGE

PROTEGE

SOURCE PROTEGEE

EAU DE PLUIE

CAMION CITERNE

CHARRETTE AVEC PETITE CITERNE

EAU EN BOUTEILLE

EAU EN SACHET

PAS PROTEGE

SOURCE NON PROTEGEE

EAU DE SURFACE (RIVIERE, BARRAGE, LAC, MARE, COURANT, CANAL, SYSTEME D’IRRIGATION)

AUTRE

Total

Pourcen-tage utilisant de l’eau de boisson de sources améliorées1

Nombre de membres des ménages

Total

39.1

2.4

0.5

1.5

1.1

2.2

1.3

11.4

2.6

11.1

24.6

0.2

0.5

0.6

0.0

0.8

100

98.0

43,638

Area

URBAIN

38.6

0.5

0.2

0.2

0.4

0.8

0.6

9.2

2.1

12.6

33.4

0.3

0.1

0.1

0.0

0.9

100

99.0

29,853

RURAL

40.0

6.6

1.1

4.5

2.6

5.3

2.7

16.4

3.7

7.7

5.4

0.0

1.5

1.7

0.1

0.7

100

95.9

13,785

Region

DISTRICT TUNIS

32.6

0.7

0.2

0.1

0.0

0.3

0.0

0.0

1.9

15.6

47.0

0.2

0.0

0.0

0.0

1.2

100

98.8

10,398

NORD EST

44.5

1.7

0.3

0.5

1.1

2.0

1.3

0.1

0.1

25.2

22.1

0.2

0.4

0.0

0.1

0.3

100

99.2

6,146

NORD OUEST

55.8

3.9

1.5

6.0

0.4

0.7

9.2

0.1

0.3

3.8

11.4

0.1

1.4

5.2

0.3

0.0

100

93.2

4,777

CENTRE EST

34.2

3.3

0.6

0.2

0.7

0.4

0.0

28.3

2.4

0.7

27.3

0.2

0.0

0.0

0.0

1.5

100

98.5

10,289

CENTRE OUEST

49.3

5.2

0.6

5.6

5.2

8.2

0.4

3.3

6.3

7.6

6.1

0.0

1.1

0.3

0.1

0.8

100

97.8

5,584

SUD EST

25.6

1.5

0.2

0.0

0.4

5.6

0.0

47.0

1.1

3.9

14.3

0.1

0.2

0.0

0.0

0.1

100

99.7

3,957

SUD OUEST

39.3

0.5

0.3

0.0

0.3

2.2

0.8

1.1

10.7

33.2

8.7

0.2

2.4

0.0

0.0

0.4

100

97.1

2,486

Education of household head

Pre-primary or none

46.6

4.7

0.9

3.1

1.3

3.0

2.6

9.4

3.9

10.6

9.8

0.2

1.4

1.5

0.1

0.8

100

96.2

7,899

Primary

43.8

3.2

0.8

1.9

1.6

3.0

1.4

13.1

2.8

12.2

14.0

0.1

0.5

0.7

0.1

0.9

100

97.9

16,191

Secondary

36.1

1.1

0.2

0.6

0.7

1.5

0.8

11.1

2.0

11.9

32.5

0.2

0.2

0.2

0.0

0.8

100

98.8

14,179

Higher

20.9

0.2

0.0

0.2

0.4

0.7

0.4

10.4

1.4

6.2

57.9

0.2

0.0

0.2

0.0

0.7

100

99.0

5,294

Missing/DK

62.5

0.0

0.0

18.1

0.0

0.0

0.0

2.1

0.0

0.0

17.4

0.0

0.0

0.0

0.0

0.0

100

100.0

75

Wealth index quintile

Poorest

28.7

8.9

1.8

7.0

4.1

7.4

4.0

20.7

3.9

6.0

1.9

0.0

2.2

2.6

0.2

0.6

100

94.3

8,728

Second

47.7

2.6

0.7

0.6

0.8

2.9

1.5

24.0

2.5

11.0

4.4

0.1

0.2

0.2

0.0

0.6

100

98.9

8,727

Middle

56.3

0.4

0.1

0.0

0.5

0.7

0.8

9.1

2.3

15.6

13.1

0.1

0.1

0.2

0.0

0.8

100

98.9

8,726

Fourth

45.2

0.1

0.0

0.0

0.1

0.1

0.0

2.7

2.7

15.3

32.2

0.4

0.0

0.0

0.0

1.2

100

98.8

8,731

Richest

17.5

0.1

0.0

0.0

0.0

0.0

0.2

0.7

1.4

7.6

71.3

0.3

0.0

0.0

0.0

0.9

100

99.1

8,725

tab.1.1 %>%   
 kable(format = "latex", digits = 1, format.args = list(big.mark = ",", scientific = FALSE)) %>%   
 kable\_styling(c("striped"), font\_size = 11) %>%   
# pack\_rows("Total", 1, 1) %>%   
 pack\_rows(label(hh$HH6), 2, 3) %>%   
 pack\_rows(label(hh$HH7), 4, 10) %>%   
 pack\_rows(label(hh$helevel), 11, 15) %>%   
 pack\_rows(label(hh$windex5), 16, 20) %>%   
 add\_header\_above(c(" " , "Eau courante" = 4, " " = 15)) %>%   
 add\_header\_above(c(" ", "Sources améliorées" = 12, "Sources non améliorées" = 4, " " = 3)) %>%   
 add\_header\_above(c(" ", "Source principale d'eau de boisson" = 19))   
  
  
a = data.frame(tab.1.1);   
  
a = lapply(a, function(x) round(x, digits = 1)) %>% data.frame(); colnames(a) = colnames(tab.1.1)  
rownames(a) = rownames(tab.1.1)  
  
aa = colnames(a)  
  
  
ft = flextable(a %>% rownames\_to\_column(" ")) %>% autofit() %>%   
 fontsize(part = "all", size = 4) %>% bold(part = "header") %>%   
 add\_header(`DANS LE LOGEMENT` = "Eau courante",  
 `DANS LA CONCESSION/JARDIN/ PARCELLE` = "Eau courante",  
 `CHEZ LE VOISIN` = "Eau courante",  
 `ROBIENT PUBLIC/BORNE FONTAINE` = "Eau courante") %>%   
 add\_header(  
 `DANS LE LOGEMENT` = "Sources améliorées",  
 `DANS LA CONCESSION/JARDIN/ PARCELLE` = "Sources améliorées",  
 `CHEZ LE VOISIN` = "Sources améliorées",  
 `ROBIENT PUBLIC/BORNE FONTAINE` = "Sources améliorées",   
 `PUITS A POMPE/FORAGE` = "Sources améliorées",  
 `PROTEGE` = "Sources améliorées",  
 `SOURCE PROTEGEE` = "Sources améliorées",  
 `EAU DE PLUIE` = "Sources améliorées",  
 `CAMION CITERNE` = "Sources améliorées",  
 `CHARRETTE AVEC PETITE CITERNE` = "Sources améliorées",  
 `EAU EN BOUTEILLE` = "Sources améliorées",  
 `EAU EN SACHET` = "Sources améliorées",  
 `PAS PROTEGE` = "Sources non améliorées",  
 `SOURCE NON PROTEGEE` = "Sources non améliorées",  
 `EAU DE SURFACE (RIVIERE, BARRAGE, LAC, MARE, COURANT, CANAL, SYSTEME D’IRRIGATION)` = "Sources non améliorées",  
 `AUTRE` = "Sources non améliorées"  
 ) %>%   
 add\_header(  
 `DANS LE LOGEMENT` = "Source principale d'eau de boisson",  
 `DANS LA CONCESSION/JARDIN/ PARCELLE` = "Source principale d'eau de boisson",  
 `CHEZ LE VOISIN` = "Source principale d'eau de boisson",  
 `ROBIENT PUBLIC/BORNE FONTAINE` = "Source principale d'eau de boisson",   
 `PUITS A POMPE/FORAGE` = "Source principale d'eau de boisson",  
 `PROTEGE` = "Source principale d'eau de boisson",  
 `SOURCE PROTEGEE` = "Source principale d'eau de boisson",  
 `EAU DE PLUIE` = "Source principale d'eau de boisson",  
 `CAMION CITERNE` = "Source principale d'eau de boisson",  
 `CHARRETTE AVEC PETITE CITERNE` = "Source principale d'eau de boisson",  
 `EAU EN BOUTEILLE` = "Source principale d'eau de boisson",  
 `EAU EN SACHET` = "Source principale d'eau de boisson",  
 `PAS PROTEGE` = "Source principale d'eau de boisson",  
 `SOURCE NON PROTEGEE` = "Source principale d'eau de boisson",  
 `EAU DE SURFACE (RIVIERE, BARRAGE, LAC, MARE, COURANT, CANAL, SYSTEME D’IRRIGATION)` = "Source principale d'eau de boisson",  
 `AUTRE` = "Source principale d'eau de boisson"  
 ) %>%   
 merge\_h(i = 1:3, part = "header") %>%   
 theme\_zebra() %>%   
 align(align = "center", part = "all") %>% fontsize(part = "header", size = 5)  
 #border\_inner\_v(part = "all", border = fp\_border(color = "gray")) %>% fix\_border\_issues()  
   
   
   
   
   
  
# add\_header(`DANS LE LOGEMENT` = "Eau courante",  
# `DANS LA CONCESSION/JARDIN/ PARCELLE` = "Eau courante",  
# `CHEZ LE VOISIN` = "Eau courante",  
# `ROBIENT PUBLIC/BORNE FONTAINE` = "Eau courante")   
  
FitFlextableToPage <- function(ft, pgwidth = 8){  
  
 ft\_out <- ft %>% autofit()  
  
 ft\_out <- width(ft\_out, width = dim(ft\_out)$widths\*pgwidth /(flextable\_dim(ft\_out)$widths))  
 return(ft\_out)  
}  
  
 FitFlextableToPage(ft, pgwidth = 8)

# 10 - WS.1.2

Reminder: change ALL case whens into nested ifelse, case\_when function just burned me on 1.2 table…

hh <- read\_sav(here("Data/Tunisia\_hh.sav"))  
  
hh <- hh %>% filter(HH46 == 1) %>% mutate(hhweightHH48 = HH48 \* hhweight) %>%   
 mutate(nhhmem = 1 \* hhweightHH48)  
  
var\_lab(hh$nhhmem) = "Number of household members"; val\_lab(hh$nhhmem) = make\_labels("  
 1  
 ")  
# reload data as start over as in SPS syntax file  
  
hh <- hh %>% mutate(time = ifelse(WS4 %in% 0:30, 2,   
 ifelse(WS4 %in% 31:990, 3,  
 ifelse(WS4 %in% c(998, 999), 9,  
 WS4)))) %>%   
 mutate(time = ifelse(WS1 %in% 11:13 |  
 WS2 %in% 11:13 |  
 WS3 %in% 1:2, 1, time))  
  
  
  
  
var\_lab(hh$time) = "Time to source of drinking water"; val\_lab(hh$time) = make\_labels("   
 1 Water on premises  
 2 Up to and including 30 minutes [A]  
 3 More than 30 minutes  
 9 DK/Missing  
 ")  
  
  
##################### BEGIN 10 - WS.SPS ################################  
  
  
hh <- hh %>%   
 mutate(drinkingWater = ifelse(WS1 %in% c(11:14, 21, 31, 41, 51, 61, 71, 72, 91, 92), 1, 2))  
  
var\_lab(hh$drinkingWater) = "Main source of drinking water"  
val\_lab(hh$drinkingWater) = make\_labels("  
 1 Improved sources  
 2 Unimproved sources  
")  
  
  
  
hh <- hh %>%   
 mutate(toiletType = case\_when(  
 WS11 %in% c(11:13, 18, 21, 22, 31) ~ 1,  
 WS11 == 95 ~ 3,  
 TRUE ~ 2  
 ))  
  
  
var\_lab(hh$toiletType) = "Type of sanitation facility"  
val\_lab(hh$toiletType) = make\_labels("  
 1 Improved  
 2 Unimproved  
 3 Open defecation (no facility, bush, field)  
 ")  
  
  
hh <- hh %>% mutate(flush = NA) %>% mutate(flush = case\_when(  
 WS11 %in% c(1:14, 18) ~ 1  
))  
  
var\_lab(hh$flush) = ""; val\_lab(hh$flush) = make\_labels("   
 1 Flush/Pour flush to:  
 ")  
  
  
  
hh <- hh %>% mutate(sharedToilet = case\_when(  
 WS17 %in% c(1:5) ~ 1,  
 WS17 %in% 97:99 ~ 9,  
 NA ~ 0,  
 TRUE ~ 2  
))  
  
hh <- hh %>% mutate(sharedToilet = ifelse(WS16 == 2, 3, sharedToilet))  
  
var\_lab(hh$sharedToilet) = " "  
val\_lab(hh$sharedToilet) = make\_labels("  
 0 Not shared  
 1 5 households or less  
 2 More than 5 households  
 3 Public facility  
 9 DK/Missing  
 ")  
  
  
##################### END 10 - WS.SPS ################################  
  
var\_lab(hh$drinkingWater) = "Time to source of drinking Water"  
val\_lab(hh$drinkingWater) = make\_labels("  
 1 Users of improved drinking water sources   
 2 Users of unimproved drinking water sources  
 ")  
  
hh <- hh %>%   
 mutate(INDWS2 = 0) %>%   
 mutate(INDWS2 = case\_when(  
 WS1 %in% c(11, 12) ~ 100,  
 (WS1 %in% c(61, 71, 72)) & (WS4 <= 30) ~ 100,  
 (WS1 %in% c(13, 14, 21, 31, 41, 51)) & (WS3 %in% 1:2 | WS4 <= 30) ~ 100,  
 (WS1 %in% 91:92) & (WS2 %in% 11:12) ~ 100,  
 (WS1 %in% 91:92) & (WS2 %in% c(61, 71, 72) & WS4 <= 30) ~ 100,  
 (WS1 %in% 91:92) & (WS2 %in% c(13, 14, 21, 31, 41, 51)) & (WS3 %in% 1:2 | WS4 <= 30) ~ 100  
 ) )  
  
hh$total = 1  
var\_lab(hh$INDWS2) = "Percentage using basic drinking water services [1]"  
var\_lab(hh$total) = "Total"; val\_lab(hh$total) = make\_labels("  
 1   
 ")

## 10.1.2 Tables

The general code process for creating the tables in R is as follows: (there may be a more clever way to do this in one table command, but this works for now)

Each xtabs command computes a contingency table We go group by group - so e.g. starting off with contingency table of HH6 (Area) vs. WS1 (drinking water source) , or for 1.2 Area vs time

This one xtabs command is a single building block in our SFR table, we then stack these downwards using the rbind command in R, which “rowbinds” blocks together.

Naturally - Using rbind to rowbind two objects together (table / dataframe / list etc) requires exact same columns in each object, similarly cbind (column binding) two objects requires exact same rows

We rowbind together downwards - HH6 - Area, HH7 Region, helevel education level, and windex5 wealth quintile. For the table here in 1.2, there’s an additional step of performing this sequence on a filtered dataframe for drinkingWater == 1, then using cbind to bind together the same table but for drinkingWater == 2

df.unweighted <- hh %>% summarize(Count = sum(HH48)) %>%   
 rbind.data.frame(hh %>% group\_by(as\_factor(HH6)) %>%   
 summarize(Count = sum(HH48)) %>% select(Count)) %>%   
 rbind.data.frame(hh %>% group\_by(as\_factor(HH7)) %>%   
 summarize(Count = sum(HH48)) %>% select(Count)) %>%   
 rbind.data.frame(hh %>% group\_by(as\_factor(helevel)) %>%   
 summarize(Count = sum(HH48)) %>% select(Count)) %>%   
 rbind.data.frame(hh %>% group\_by(as\_factor(windex5)) %>%   
 summarize(Count = sum(HH48)) %>% select(Count));   
  
df.unweighted$Count = round(df.unweighted$Count);   
rownames(df.unweighted) = c("Total", levels(as\_factor(hh$HH6)), levels(as\_factor(hh$HH7)), levels(as\_factor(hh$helevel)),   
 levels(as\_factor(hh$windex5)))

## Warning: Setting row names on a tibble is deprecated.

# to be turned into function - this is likely common task in every chapter...  
N <- hh %>% summarize(Count = sum(nhhmem)) %>%   
 rbind.data.frame(hh %>% group\_by(as\_factor(HH6)) %>%   
 summarize(Count = sum(nhhmem)) %>% select(Count)) %>%   
 rbind.data.frame(hh %>% group\_by(as\_factor(HH7)) %>%   
 summarize(Count = sum(nhhmem)) %>% select(Count)) %>%   
 rbind.data.frame(hh %>% group\_by(as\_factor(helevel)) %>%   
 summarize(Count = sum(nhhmem)) %>% select(Count)) %>%   
 rbind.data.frame(hh %>% group\_by(as\_factor(windex5)) %>%   
 summarize(Count = sum(nhhmem)) %>% select(Count));   
  
N$Count = round(N$Count)  
N <- N %>% mutate(Total = Count)  
N <- data.frame(N); colnames(N)[1] = c("Number of household members")  
rownames(N) = c("Total", levels(as\_factor(hh$HH6)), levels(as\_factor(hh$HH7)), levels(as\_factor(hh$helevel)),   
 levels(as\_factor(hh$windex5)))  
  
  
  
  
tab.improved = xtabs(nhhmem ~ as\_factor(time), hh %>% filter(drinkingWater == 1)) %>%   
 rbind(xtabs(nhhmem ~ as\_factor(HH6) + as\_factor(time), hh %>% filter(drinkingWater == 1))) %>%   
 rbind(xtabs(nhhmem ~ as\_factor(HH7) + as\_factor(time), hh %>% filter(drinkingWater == 1))) %>%  
 rbind(xtabs(nhhmem ~ as\_factor(helevel) + as\_factor(time), hh %>% filter(drinkingWater == 1))) %>%   
 rbind(xtabs(nhhmem ~ as\_factor(windex5) + as\_factor(time), hh %>% filter(drinkingWater == 1)))  
  
  
tab.unimproved = xtabs(nhhmem ~ as\_factor(time), hh %>% filter(drinkingWater == 2)) %>%   
 rbind(xtabs(nhhmem ~ as\_factor(HH6) + as\_factor(time), hh %>% filter(drinkingWater == 2))) %>%   
 rbind(xtabs(nhhmem ~ as\_factor(HH7) + as\_factor(time), hh %>% filter(drinkingWater == 2))) %>%  
 rbind(xtabs(nhhmem ~ as\_factor(helevel) + as\_factor(time), hh %>% filter(drinkingWater == 2))) %>%   
 rbind(xtabs(nhhmem ~ as\_factor(windex5) + as\_factor(time), hh %>% filter(drinkingWater == 2)))  
  
  
tab.basic = xtabs(nhhmem ~ as\_factor(INDWS2), hh) %>%   
 rbind(xtabs(nhhmem ~ as\_factor(HH6) + as\_factor(INDWS2), hh)) %>%   
 rbind(xtabs(nhhmem ~ as\_factor(HH7) + as\_factor(INDWS2), hh)) %>%  
 rbind(xtabs(nhhmem ~ as\_factor(helevel) + as\_factor(INDWS2), hh)) %>%   
 rbind(xtabs(nhhmem ~ as\_factor(windex5) + as\_factor(INDWS2), hh))  
  
  
tab.1.2 = cbind(tab.improved, tab.unimproved, N$Total, tab.basic, N$`Number of household members`)  
  
  
  
tab.1.2 = apply(tab.1.2[ , 1:(ncol(tab.1.2)-1)], 2, function(x) (x \* 100) / tab.1.2[ , ncol(tab.1.2) - 2]) %>% cbind(N$`Number of household members`)  
  
colnames(tab.1.2)[(ncol(tab.1.2) - 2):ncol(tab.1.2)] = c("Total",  
 "Percentage using basic drinking water services [1]",  
 "Number of household members")  
  
  
  
rownames(tab.1.2)[1] = "Total"

### parenth function

# inputs - df is the dataframe in which we want to suppress values / apply rule  
# N is the vector of unweighted samples  
  
samp\_check\_fun <- function(df, N, inds) {  
 ind = ifelse(N < 25, 1,   
 ifelse(N %in% 25:49, 2,  
 3))  
   
   
   
 df[ind == 1, inds] = "\*"  
 df[ind == 2, inds] =   
   
   
}

# SFR Table Ch10 WS 1.2

header1 = label(hh$time)  
  
tab.1.2 %>%   
 kable(digits = 1, format.args = list(big.mark = ",", scientific = FALSE)) %>%   
 kable\_styling(c("striped"), font\_size = 12) %>%   
 pack\_rows(label(hh$HH6), 2, 3) %>%   
 pack\_rows(label(hh$HH7), 4, 10) %>%   
 pack\_rows(label(hh$helevel), 11, 15) %>%   
 pack\_rows(label(hh$windex5), 16, 20) %>%   
 add\_header\_above(c(" ", "Users of improved drinking water sources" = 4,   
 "Users of unimproved drinking water sources" = 4,  
 " " = 3)) %>%   
 add\_header\_above(c(" ", "Time to source of drinking water" = 8, " " = 3))

Time to source of drinking water

Users of improved drinking water sources

Users of unimproved drinking water sources

Water on premises

Up to and including 30 minutes [A]

More than 30 minutes

DK/Missing

Water on premises

Up to and including 30 minutes [A]

More than 30 minutes

DK/Missing

Total

Percentage using basic drinking water services [1]

Number of household members

Total

86.6

9.1

2.1

0.1

1.0

0.6

0.4

0.0

100

85.9

43,638

Area

URBAIN

91.4

6.8

0.7

0.1

0.9

0.2

0.0

0.0

100

86.4

29,853

RURAL

76.3

14.1

5.3

0.3

1.4

1.4

1.1

0.1

100

84.6

13,785

Region

DISTRICT TUNIS

94.5

4.3

0.0

0.0

1.2

0.0

0.0

0.0

100

84.4

10,398

NORD EST

94.7

3.8

0.6

0.1

0.4

0.3

0.1

0.0

100

74.8

6,146

NORD OUEST

77.6

10.5

5.1

0.0

0.6

3.7

2.3

0.2

100

85.0

4,777

CENTRE EST

86.1

11.3

0.9

0.2

1.5

0.0

0.0

0.0

100

96.8

10,289

CENTRE OUEST

71.5

17.1

9.1

0.0

1.0

0.7

0.5

0.0

100

83.7

5,584

SUD EST

90.6

7.5

0.9

0.7

0.1

0.0

0.2

0.0

100

95.3

3,957

SUD OUEST

81.0

15.3

0.4

0.5

2.4

0.3

0.1

0.1

100

65.2

2,486

Education of household head

Pre-primary or none

81.4

10.5

4.2

0.1

1.4

1.5

0.8

0.2

100

82.8

7,899

Primary

84.5

10.9

2.3

0.2

1.0

0.6

0.5

0.0

100

84.7

16,191

Secondary

89.8

7.5

1.2

0.2

1.0

0.1

0.0

0.0

100

86.6

14,179

Higher

92.3

5.8

0.9

0.0

0.7

0.2

0.0

0.0

100

92.1

5,294

Missing/DK

82.3

12.2

5.9

0.0

0.0

0.0

0.0

0.0

100

94.5

75

Wealth index quintile

Poorest

67.6

18.8

7.5

0.3

1.7

2.2

1.7

0.1

100

81.8

8,728

Second

84.4

12.4

1.9

0.2

0.6

0.4

0.0

0.0

100

88.0

8,727

Middle

89.9

7.8

1.0

0.2

0.9

0.3

0.0

0.0

100

83.8

8,726

Fourth

94.0

4.6

0.2

0.0

1.2

0.0

0.0

0.0

100

84.3

8,731

Richest

97.1

2.0

0.0

0.0

0.9

0.0

0.0

0.0

100

91.5

8,725

a = data.frame(tab.1.2);   
  
a = lapply(a, function(x) round(x, digits = 1)) %>% data.frame(); colnames(a) = colnames(tab.1.2)  
rownames(a) = rownames(tab.1.2)  
colnames(a)[5:8] = c(" Water on premises", " Up to and including 30 minutes [A]",  
 " More than 30 minutes", " DK/Missing")  
aa = colnames(a)  
  
  
ft = flextable(a %>% rownames\_to\_column(" ")) %>% autofit() %>%   
 fontsize(part = "all", size = 4) %>% bold(part = "header") %>%   
 add\_header(`Water on premises` = "Users of improved drinking water sources",  
 `Up to and including 30 minutes [A]` = "Users of improved drinking water sources",  
 `More than 30 minutes` = "Users of improved drinking water sources",  
 `DK/Missing` = "Users of improved drinking water sources",  
 ` Water on premises` = "Users of unimproved drinking water sources",  
 ` Up to and including 30 minutes [A]` = "Users of unimproved drinking water sources",  
 ` More than 30 minutes` = "Users of unimproved drinking water sources",  
 ` DK/Missing` = "Users of unimproved drinking water sources") %>%   
add\_header(  
 `Water on premises` = "Time to source of drinking water",  
 `Up to and including 30 minutes [A]` = "Time to source of drinking water",  
 `More than 30 minutes` = "Time to source of drinking water",  
 `DK/Missing` = "Time to source of drinking water",  
 ` Water on premises` = "Time to source of drinking water",  
 ` Up to and including 30 minutes [A]` = "Time to source of drinking water",  
 ` More than 30 minutes` = "Time to source of drinking water",  
 ` DK/Missing` = "Time to source of drinking water"  
) %>%   
 merge\_h(i = 1:3, part = "header") %>%   
 theme\_zebra() %>%   
 align(align = "center", part = "all") %>% fontsize(part = "header", size = 5)  
 #border\_inner\_v(part = "all", border = fp\_border(color = "gray")) %>% fix\_border\_issues()  
   
   
   
   
   
  
# add\_header(`DANS LE LOGEMENT` = "Eau courante",  
# `DANS LA CONCESSION/JARDIN/ PARCELLE` = "Eau courante",  
# `CHEZ LE VOISIN` = "Eau courante",  
# `ROBIENT PUBLIC/BORNE FONTAINE` = "Eau courante")   
  
FitFlextableToPage <- function(ft, pgwidth = 8){  
  
 ft\_out <- ft %>% autofit()  
  
 ft\_out <- width(ft\_out, width = dim(ft\_out)$widths\*pgwidth /(flextable\_dim(ft\_out)$widths))  
 return(ft\_out)  
}  
  
 FitFlextableToPage(ft, pgwidth = 8)

|  | **Time to source of drinking water** | | | | | | | |  | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Users of improved drinking water sources** | | | | **Users of unimproved drinking water sources** | | | |  | | |
|  | **Water on premises** | **Up to and including 30 minutes [A]** | **More than 30 minutes** | **DK/Missing** | **Water on premises** | **Up to and including 30 minutes [A]** | **More than 30 minutes** | **DK/Missing** | **Total** | **Percentage using basic drinking water services [1]** | **Number of household members** |
| Total | 86.6 | 9.1 | 2.1 | 0.1 | 1.0 | 0.6 | 0.4 | 0.0 | 100 | 85.9 | 43638 |
| URBAIN | 91.4 | 6.8 | 0.7 | 0.1 | 0.9 | 0.2 | 0.0 | 0.0 | 100 | 86.4 | 29853 |
| RURAL | 76.3 | 14.1 | 5.3 | 0.3 | 1.4 | 1.4 | 1.1 | 0.1 | 100 | 84.6 | 13785 |
| DISTRICT TUNIS | 94.5 | 4.3 | 0.0 | 0.0 | 1.2 | 0.0 | 0.0 | 0.0 | 100 | 84.4 | 10398 |
| NORD EST | 94.7 | 3.8 | 0.6 | 0.1 | 0.4 | 0.3 | 0.1 | 0.0 | 100 | 74.8 | 6146 |
| NORD OUEST | 77.6 | 10.5 | 5.1 | 0.0 | 0.6 | 3.7 | 2.3 | 0.2 | 100 | 85.0 | 4777 |
| CENTRE EST | 86.1 | 11.3 | 0.9 | 0.2 | 1.5 | 0.0 | 0.0 | 0.0 | 100 | 96.8 | 10289 |
| CENTRE OUEST | 71.5 | 17.1 | 9.1 | 0.0 | 1.0 | 0.7 | 0.5 | 0.0 | 100 | 83.7 | 5584 |
| SUD EST | 90.6 | 7.5 | 0.9 | 0.7 | 0.1 | 0.0 | 0.2 | 0.0 | 100 | 95.3 | 3957 |
| SUD OUEST | 81.0 | 15.3 | 0.4 | 0.5 | 2.4 | 0.3 | 0.1 | 0.1 | 100 | 65.2 | 2486 |
| Pre-primary or none | 81.4 | 10.5 | 4.2 | 0.1 | 1.4 | 1.5 | 0.8 | 0.2 | 100 | 82.8 | 7899 |
| Primary | 84.5 | 10.9 | 2.3 | 0.2 | 1.0 | 0.6 | 0.5 | 0.0 | 100 | 84.7 | 16191 |
| Secondary | 89.8 | 7.5 | 1.2 | 0.2 | 1.0 | 0.1 | 0.0 | 0.0 | 100 | 86.6 | 14179 |
| Higher | 92.3 | 5.8 | 0.9 | 0.0 | 0.7 | 0.2 | 0.0 | 0.0 | 100 | 92.1 | 5294 |
| Missing/DK | 82.3 | 12.2 | 5.9 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 100 | 94.5 | 75 |
| Poorest | 67.6 | 18.8 | 7.5 | 0.3 | 1.7 | 2.2 | 1.7 | 0.1 | 100 | 81.8 | 8728 |
| Second | 84.4 | 12.4 | 1.9 | 0.2 | 0.6 | 0.4 | 0.0 | 0.0 | 100 | 88.0 | 8727 |
| Middle | 89.9 | 7.8 | 1.0 | 0.2 | 0.9 | 0.3 | 0.0 | 0.0 | 100 | 83.8 | 8726 |
| Fourth | 94.0 | 4.6 | 0.2 | 0.0 | 1.2 | 0.0 | 0.0 | 0.0 | 100 | 84.3 | 8731 |
| Richest | 97.1 | 2.0 | 0.0 | 0.0 | 0.9 | 0.0 | 0.0 | 0.0 | 100 | 91.5 | 8725 |

# flextables experimentation

a = data.frame(tab.1.1);   
  
a = lapply(a, function(x) round(x, digits = 1)) %>% data.frame(); colnames(a) = colnames(tab.1.1)  
rownames(a) = rownames(tab.1.1)  
  
flextable(a %>% rownames\_to\_column("a"), col\_keys = names(a)) %>% autofit() %>%   
 add\_header\_lines("Users of improved drinking water sources") %>%   
 fontsize(part = "header", size = 14) %>% bold(part = "header") %>%   
 border\_inner\_v(border = fp\_border(color = "gray"), part = "all")  
  
  
aa = colnames(a)  
  
flextable(a %>% rownames\_to\_column(" ")) %>% autofit() %>%   
 fontsize(part = "header", size = 10) %>% bold(part = "header") %>%   
 fontsize(part = "body", size = 7) %>%   
 add\_header(`DANS LE LOGEMENT` = "Eau courante") %>%   
 merge\_at(i = 1, j = 2:4, part = "header") %>%   
 add\_header(`PROTEGE` = "Sources améliorées",  
 `AUTRE` = "Sources non améliorées") %>%   
 # merge\_at(i = 1, j = 5:12, part = "header") %>%   
 add\_header(`CAMION CITERNE` = "Source principale d'eau de boisson") %>%   
 border\_inner\_v(border = fp\_border(color = "gray")) %>% theme\_zebra()

flextable(a) %>% set\_header\_df(mapping = b, key = "col\_keys")  
  
ft <- flextable( head( iris ) ); ft  
ft <- add\_footer\_lines(ft, values = "blah blah"); ft  
ft <- add\_footer\_lines(ft, values = c("blah 1", "blah 2"))  
autofit(ft)  
ft <- flextable( head( iris ) ); ft  
ft <- add\_header\_lines(ft, values = "blah blah"); ft  
ft <- add\_header\_lines(ft, values = c("blah 1", "blah 2")); ft  
autofit(ft)

atypology <- data.frame(  
 col\_keys = colnames(a),  
 what = c(rep("a", 4), rep(" ", 15)), stringsAsFactors = FALSE)  
  
  
typology <- data.frame(  
 col\_keys = c( "Sepal.Length", "Sepal.Width", "Petal.Length",  
 "Petal.Width", "Species" ),  
 what = c("Sepal", "Sepal", "Petal", "Petal", "Species"),  
 measure = c("Length", "Width", "Length", "Width", "Species"),  
 stringsAsFactors = FALSE )  
  
ft <- flextable( head( iris ))  
ft <- set\_header\_df(ft, mapping = typology, key = "col\_keys" )  
ft <- merge\_h(ft, part = "header")  
ft <- merge\_v(ft, j = "Species", part = "header")  
ft <- theme\_vanilla(ft)  
ft  
  
typology <- data.frame(  
 col\_keys = c( "Sepal.Length", "Sepal.Width", "Petal.Length",  
 "Petal.Width", "Species" ),  
 unit = c("(cm)", "(cm)", "(cm)", "(cm)", ""),  
 stringsAsFactors = FALSE )  
ft <- set\_footer\_df(ft, mapping = typology, key = "col\_keys" )  
ft <- italic(ft, italic = TRUE, part = "footer" )  
ft <- theme\_booktabs(ft)  
ft