

Exercice 2.1: Accès à l'information et attitudes face à la violence conjugale

Miranda Melson

08 June 2021

Dans cet exercice, vous aller travailler à partir de la base de donnée `dhs_ipv` qui a été utilisée dans cette étude :

Pierotti, Rachel. (2013). "Increasing Rejection of Intimate Partner Violence: Evidence of Global Cultural Diffusion." *American Sociological Review*, 78: 240-265.

Il s'agit d'une base de données dont l'unité d'analyse sont les pays sur lesquels on dispose d'un certain nombre d'information. Ces informations sont présentées dans le tableau ci-dessous:

Name	Description
<code>beat_goesout</code>	Pourcentage de femmes dans chaque pays qui pensent qu'un mari est justifié de battre sa femme si elle sort sans le lui dire.
<code>beat_burnfood</code>	Pourcentage de femmes dans chaque pays qui pensent qu'un mari a le droit de battre sa femme si elle brûle sa nourriture.
<code>no_media</code>	Pourcentage de femmes dans chaque pays qui ont rarement accès un journal, une radio ou une télévision.
<code>sec_school</code>	Pourcentage de femmes dans chaque pays ayant un niveau d'éducation secondaire ou supérieur.
<code>year</code>	Année de l'enquête
<code>region</code>	Région du monde
<code>country</code>	pays

Name	Description
<code>beat_goesout</code>	Percentage of women in each country that think a husband is justified to beat his wife if she goes out with him.
<code>beat_burnfood</code>	Percentage of women in each country that think a husband is justified to beat his wife if she burns his food.
<code>no_media</code>	Percentage of women in each country that rarely encounter a newspaper, radio, or television.
<code>sec_school</code>	Percentage of women in each country with secondary or higher education.
<code>year</code>	Year of the survey
<code>region</code>	Region of the world
<code>country</code>	Country

Notez qu'il existe dans ce fichier de données, deux indicateurs/variables qui mesurent les *attitudes envers la violence domestique*: `beat_goesout` et `beat_burnfood`. Ce sont ces variables que nous voulons expliquer/comprendre. On les appelle des variables dépendantes ou variables à expliquer ou des *outcomes*. Il existe également deux indicateurs/variables qui mesurent l'*accès à l'information*: `sec_school` et `no_media`. Celles-ci sont appelées des variables explicatives.

Comme toujours, il faut prévoir un premier *chunk* où vous installez vos packages, les chargez et chargez la base de données. C'est une procédure qu'il faut toujours suivre. Sachez aussi que vous devez ouvrir ce

fichier RMarkdown et travailler directement dedans.

#1. Effacer l'environnement

```
rm(list = ls())      # Permet d'effacer l'environnement
```

#2. Installer de nouveaux packages dont vous avez besoins

#3. Chargez les packages

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.0.5
```

```
## -- Attaching packages ----- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.3      v purrr  0.3.4
```

```
## v tibble  3.1.2      v dplyr  1.0.6
```

```
## v tidyr   1.1.3      v stringr 1.4.0
```

```
## v readr   1.4.0      v forcats 0.5.1
```

```
## Warning: package 'ggplot2' was built under R version 4.0.5
```

```
## Warning: package 'tibble' was built under R version 4.0.5
```

```
## Warning: package 'tidyr' was built under R version 4.0.5
```

```
## Warning: package 'dplyr' was built under R version 4.0.5
```

```
## Warning: package 'forcats' was built under R version 4.0.5
```

```
## -- Conflicts ----- tidyverse_conflicts() --
```

```
## x dplyr::filter() masks stats::filter()
```

```
## x dplyr::lag()     masks stats::lag()
```

#4. Ouvrir votre base de données

```
dhs <- read_csv("../Données/dhs_ipv.csv")
```

```
## Warning: Missing column names filled in: 'X1' [1]
```

```
##
```

```
## -- Column specification -----
```

```
## cols(
```

```
##   X1 = col_double(),
```

```
##   beat_burnfood = col_double(),
```

```
##   beat_goesout = col_double(),
```

```
##   sec_school = col_double(),
```

```
##   no_media = col_double(),
```

```
##   country = col_character(),
```

```
##   year = col_double(),
```

```
##   region = col_character()
```

```
## )
```

Comme vous avez pu le voir en classe, quand vous chargez Tidyverse, il charge l'ensemble des packages qui sont inclus dedans. C'est la dernière fois que je dresse la table pour vous :)

PARTIE A: Sélection des variables et des observations

Question 1

Votre base de données comprend combien d'observation et combien de variables ? Soyez **concis** dans votre réponse.

Votre réponse

```
dim(dhs)
```

```
## [1] 151 8
```

Question 2

Dites le type de chaque variable de la base de données. Comment est ce que R comprend ce type de variable? Présentez l'information dans un tableau.

Votre réponse

Variables	Type (statistique)	Type (R)
beat_goesout	percent	numeric
beat_burnfood	percent	numeric
X	person ID	integer
sec_school	percent	numeric
no_media	percent	numeric
country	country name	character
year	number in year(s)	integer
region	continent name	character

Question 3

Quelle est le type de chaque variable de la base de données? Utiliser la fonction **class** pour le savoir. Maintenant, vous allez vous poser la question de savoir si la réponse que vous obtenez est réellement le type de chaque variable.

Votre réponse

```
class(dhs$X)
```

```
## Warning: Unknown or uninitialised column: 'X'.
```

```
## [1] "NULL"
```

```
class(dhs$beat_burnfood)
```

```
## [1] "numeric"
```

```
class(dhs$beat_goesout)
```

```
## [1] "numeric"
```

```
class(dhs$sec_school)
```

```
## [1] "numeric"
```

```
class(dhs$no_media)
```

```
## [1] "numeric"
```

```
class(dhs$country)
```

```
## [1] "character"
```

```
class(dhs$year)
```

```
## [1] "numeric"
```

```
class(dhs$region)
```

```
## [1] "character"
```

Question 4

Je vous demande de créer 5 nouvelles bases de données avec les conditions suivantes. Dans chaque cas, décrivez cette base de données en terme de population et de variables.

1. **dhs1**: beat_goesout est plus grand ou égale à 14 (ceci signifie que la valeur de la variable beat_goesout est ≥ 14)
2. **dhs2**: beat_goesout est plus grand que 13 et beat_burnfood est plus petit que 22
3. **dhs3**: sec_school plus grand que 22 et beat_goesout plus grand que 25
4. **dhs4**: comprend les pays d'Afrique sub-Saharienne dont beat_goesout est plus grand que 18 et sec_school est plus grand que 10
5. **dhs5**: beat_goesout plus petit que 15 ou beat_goesout plus grand que 22

Votre réponse

```
dhs1 <- dhs %>% filter(beat_goesout >= 14) # When 14% or greater of women think that it's justified to
dhs2 <- dhs %>% filter(beat_goesout > 13 & beat_burnfood < 22) # When more than 13% of women think that
dhs3 <- dhs %>% filter(sec_school > 22 & beat_goesout > 25) # When more than 22% have an education of m
dhs4 <- dhs %>% filter(region == "Sub-Saharan Africa" & beat_goesout >18 & sec_school > 10) # When more
dhs5 <- dhs %>% filter(beat_goesout < 15 | beat_goesout > 22) # When less than 15% or more than 22% of
```

Question 5

Créer des bases de données remplissant les conditions suivantes:

1. **dhs6**: comprend les pays dont le nom commence par c (bonus)
2. **dhs7**: comprend les pays dont le nom finit par i
3. **dhs8**: comprend tous les pays pour lequel `sec_school` est plus grand que la valeur moyenne de `sec_school`
4. **dhs9**: sélectionner les pays qui ont au moins deux années d'observation.
5. **dhs10**: comprend uniquement les données de la dernière année d'enquête de chaque pays.

Votre réponse

```
dhs6 <- dhs %>%
  filter(country == "Cambodia" | country == "Cameroon" | country == "Chad" | country == "Colombia" |
    country == "Comoros" | country == "Congo (Brazzaville)" | country == "Congo Democratic Republic" | co

dhs7 <- dhs %>%
  filter(country == "Burundi" | country == "Haiti" | country == "Malawi" | country == "Mali")

summary(dhs$sec_school)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's
##      3.10   10.18   22.40   24.40   34.90   74.60     3
```

```
dhs8 <- dhs %>%
  filter(sec_school > 3.1)

dhs9 <- dhs %>%
  select(country, year) %>%
  filter(year >= min(dhs$year)+2)

dhs10 <- dhs %>%
  group_by(country) %>%
  filter(year == max(year))
```

Partie B: Création de variables

Question 1

Quand on crée de nouvelles variables, la règle demande d'ajouter **toujours** cette nouvelle variable dans la base de donnée qu'on a utilisée pour la créer. Dans la base de données **dhs**, créer les variables suivantes:

1. **Pays_riche** qui est **TRUE** si **sec_school** est plus grand que 20 et **no_media** plus petit que 5 et **FALSE** dans le cas contraire
2. **beat_goesout_3** qui remplit les conditions suivantes:
 - elle prend la valeur **1** si **beat_goesout** est ≤ 10
 - elle prend la valeur **2** si **beat_goesout** est > 10 et ≤ 20
 - elle prend la valeur **3** si **beat_goesout** > 20

Votre réponse

```
dhs <- dhs %>%
  mutate(pays_riche = if_else(sec_school > 20 & no_media < 5, T, F)) %>%
  mutate(beat_goesout_3 = case_when(
    beat_goesout <= 10 ~ 1,
    beat_goesout > 10 & beat_goesout <= 20 ~ 2,
    beat_goesout > 20 ~ 3
  ))
```

Question 2

Sur la base des valeurs de la variable **beat_burnfood**, créer une nouvelles variable **factorielle** **beat_burnfood_cat** qui regroupe les observations en quatre catégories:

1. **Très moderne**, pays où les valeurs de **beat_burnfood** sont inférieures à 5%;
2. **Moderne**, pays où les valeurs de **beat_burnfood** sont supérieures à 5% mais inférieures à 10%;
3. **Traditionnelle**, pays où les valeurs de **beat_burnfood** sont supérieures à 10% mais inférieures à 20%;
4. **Très traditionnelle**, pays où les valeurs de **beat_burnfood** sont supérieures à 20%.

Réponse 2

```
dhs <- dhs %>%
  mutate(beat_burnfood_cat = factor(case_when(
    beat_burnfood < 5 ~ "tres moderne",
    beat_burnfood > 5 & beat_burnfood < 10 ~ "moderne",
    beat_burnfood > 10 & beat_burnfood < 20 ~ "traditionnelle",
    beat_burnfood > 20 ~ "tres traditionnelle")))

class(dhs$beat_burnfood_cat)
```

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

Partie C: Statistique univariée

Question 1

Calculer les paramètres de **position** de la variable `beat-goesout`

```
summary(dhs$beat_goesout)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's  
##      0.30  11.85   28.10   28.60  42.08   82.70    27
```

Question 4

Faite le même calcul selon les régions d'études. Commentez les résultats. PS. Vous ne devez ajouter qu'une seule ligne de commande à votre code de la question 4 pour répondre à la question.

```
tapply(dhs$beat_goesout, dhs$country, summary)
```

```
## $Albania  
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.  
##      18.6   18.6   18.6   18.6   18.6   18.6  
##  
## $Armenia  
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.  
##       3.1    6.7   10.3   11.1   15.1   19.9  
##  
## $Azerbaijan  
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.  
##      42.5   42.5   42.5   42.5   42.5   42.5  
##  
## $Bangladesh  
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's  
##     17.30  17.45   17.60   17.60  17.75   17.90     1  
##  
## $Benin  
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.  
##       7.70  22.20   36.70   29.47  40.35   44.00  
##  
## $Bolivia  
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.  
##     5.900  6.675   7.450   7.450   8.225   9.000  
##  
## $'Burkina Faso'  
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.  
##     30.30  36.05   41.80   41.80  47.55   53.30  
##  
## $Burundi  
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's  
##     48.8   48.8   48.8   48.8   48.8   48.8     1  
##  
## $Cambodia
```

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
##	8.70	19.25	29.80	23.27	30.55	31.30	1
##							
##	\$Cameroon						
##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	
##	26.50	28.27	30.05	30.05	31.82	33.60	
##							
##	\$Chad						
##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
##	NA	NA	NA	NaN	NA	NA	1
##							
##	\$Colombia						
##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
##	0.3	0.3	0.3	0.3	0.3	0.3	2
##							
##	\$Comoros						
##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	
##	28.4	28.4	28.4	28.4	28.4	28.4	
##							
##	\$'Congo (Brazzaville)'						
##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
##	40.80	43.48	46.15	46.15	48.83	51.50	1
##							
##	\$'Congo Democratic Republic'						
##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	
##	49.10	50.15	51.20	51.20	52.25	53.30	
##							
##	\$'Cote d'Ivoire'						
##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
##	26.6	26.6	26.6	26.6	26.6	26.6	1
##							
##	\$'Dominican Republic'						
##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	
##	0.800	1.000	1.200	1.733	2.200	3.200	
##							
##	\$Egypt						
##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
##	25.50	28.50	31.50	32.47	35.95	40.40	2
##							
##	\$Eritrea						
##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	
##	51.7	51.7	51.7	51.7	51.7	51.7	
##							
##	\$Ethiopia						
##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	
##	43.20	49.70	56.20	54.53	60.20	64.20	
##							
##	\$Gabon						
##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
##	19	19	19	19	19	19	1
##							
##	\$Gambia						
##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	
##	42	42	42	42	42	42	


```

##
## $Ghana
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   22.30  25.23   28.15   28.15  31.07   34.00
##
## $Guinea
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   72.40  74.97   77.55   77.55  80.12   82.70
##
## $Guyana
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##    5.8     5.8     5.8     5.8     5.8     5.8      1
##
## $Haiti
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   10.80  16.60   22.40   20.67  25.60   28.80
##
## $Honduras
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    4.300  4.725   5.150   5.150  5.575   6.000
##
## $India
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    29      29      29      29      29      29
##
## $Indonesia
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   18.20  20.95   23.70   21.97  23.85   24.00
##
## $Jordan
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   12.30  18.20   24.10   23.77  29.50   34.90      1
##
## $Kenya
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   30.70  32.88   35.05   35.05  37.23   39.40
##
## $'Kyrgyz Republic'
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   23.1    23.1    23.1    23.1    23.1    23.1
##
## $Lesotho
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   14.20  16.75   19.30   19.30  21.85   24.40
##
## $Liberia
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   28.50  31.85   35.20   35.20  38.55   41.90
##
## $Madagascar
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   14.4    15.6    16.8    16.8    18.0    19.2
##
## $Malawi

```

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	
##	5.40	9.65	13.90	11.97	15.25	16.60	
##							
##	\$Maldives						
##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	
##	13	13	13	13	13	13	
##							
##	\$Mali						
##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	
##	55.00	57.65	60.30	63.40	67.60	74.90	
##							
##	\$Mauritania						
##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
##	NA	NA	NA	NaN	NA	NA	1
##							
##	\$Moldova						
##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	
##	7.3	7.3	7.3	7.3	7.3	7.3	
##							
##	\$Morocco						
##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	
##	49.5	49.5	49.5	49.5	49.5	49.5	
##							
##	\$Mozambique						
##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
##	8.60	15.65	22.70	22.70	29.75	36.80	1
##							
##	\$Namibia						
##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
##	12.9	14.4	15.9	15.9	17.4	18.9	1
##							
##	\$Nepal						
##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	
##	0.500	4.650	8.800	7.167	10.500	12.200	
##							
##	\$Nicaragua						
##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	
##	5.8	5.8	5.8	5.8	5.8	5.8	
##							
##	\$Niger						
##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	
##	42.70	46.15	49.60	49.60	53.05	56.50	
##							
##	\$Nigeria						
##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
##	25.30	28.75	32.20	36.77	42.50	52.80	1
##							
##	\$Pakistan						
##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
##	29.6	29.6	29.6	29.6	29.6	29.6	1
##							
##	\$Peru						
##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
##	1.100	1.100	1.300	1.367	1.575	1.800	1

```

##
## $Philippines
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   4.300   4.700   5.100   6.067   6.950   8.800
##
## $Rwanda
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   26.30   31.05   35.80   32.90   36.20   36.60
##
## $'Sao Tome and Principe'
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   10.2    10.2    10.2    10.2    10.2    10.2
##
## $Senegal
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   39.90   40.65   43.55   44.17   47.08   49.70
##
## $'Sierra Leone'
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   49.7    50.5    51.3    51.3    52.1    52.9
##
## $Tajikistan
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   50.5    50.5    50.5    50.5    50.5    50.5
##
## $Tanzania
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   36.8    38.3    39.8    39.8    41.3    42.8        3
##
## $'Timor-Leste'
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   72.2    72.2    72.2    72.2    72.2    72.2
##
## $Togo
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   17.8    17.8    17.8    17.8    17.8    17.8
##
## $Turkmenistan
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   40.3    40.3    40.3    40.3    40.3    40.3
##
## $Uganda
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   37.70   44.95   52.20   48.73   54.25   56.30        3
##
## $Ukraine
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.4    0.4    0.4    0.4    0.4    0.4
##
## $Vietnam
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   NA      NA      NA      NaN      NA      NA        2
##
## $Yemen

```

```
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##      35.5   35.5    35.5    35.5   35.5    35.5
##
## $Zambia
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##      29.70   36.00   42.30   50.27   60.55   78.80
##
## $Zimbabwe
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##      22.30   25.05   27.80   27.70   30.40   33.00
```

Colombia has the minimum percentage of women who believe it's justified for their husbands to beat th

Question 5

Calculer les paramètres de **dispersion** de la variable `beat-goesout`

```
var(dhs$beat_goesout, na.rm = T)
```

```
## [1] 385.8955
```

```
sd(dhs$beat_goesout, na.rm = T)
```

```
## [1] 19.64422
```

```
range(dhs$beat_goesout, na.rm = T)
```

```
## [1] 0.3 82.7
```

Question 6

Faite le même calcul selon les régions d'études. Commentez les résultats. PS. Vous ne devez ajouter qu'une seule ligne de commande à votre code de la question 4 pour répondre à la question.

```
tapply(dhs$beat_goesout, dhs$country, var)
```

```
##              Albania              Armenia              Azerbaijan
##              NA              71.040000              NA
##      Bangladesh              Benin              Bolivia
##              NA              368.663333              4.805000
##      Burkina Faso              Burundi              Cambodia
##      264.500000              NA              NA
##      Cameroon              Chad              Colombia
##      25.205000              NA              NA
##      Comoros              Congo (Brazzaville) Congo Democratic Republic
##      NA              NA              8.820000
##      Cote d'Ivoire              Dominican Republic              Egypt
##      NA              1.653333              NA
##      Eritrea              Ethiopia              Gabon
```

##	NA	112.333333	NA
##	Gambia	Ghana	Guinea
##	NA	68.445000	53.045000
##	Guyana	Haiti	Honduras
##	NA	83.253333	1.445000
##	India	Indonesia	Jordan
##	NA	10.663333	NA
##	Kenya	Kyrgyz Republic	Lesotho
##	37.845000	NA	52.020000
##	Liberia	Madagascar	Malawi
##	89.780000	11.520000	34.163333
##	Maldives	Mali	Mauritania
##	NA	106.210000	NA
##	Moldova	Morocco	Mozambique
##	NA	NA	NA
##	Namibia	Nepal	Nicaragua
##	NA	36.223333	NA
##	Niger	Nigeria	Pakistan
##	95.220000	NA	NA
##	Peru	Philippines	Rwanda
##	NA	5.763333	32.830000
##	Sao Tome and Principe	Senegal	Sierra Leone
##	NA	21.209167	5.120000
##	Tajikistan	Tanzania	Timor-Leste
##	NA	NA	NA
##	Togo	Turkmenistan	Uganda
##	NA	NA	NA
##	Ukraine	Vietnam	Yemen
##	NA	NA	NA
##	Zambia	Zimbabwe	
##	650.303333	28.630000	

```
tapply(dhs$beat_goesout, dhs$country, sd)
```

##	Albania	Armenia	Azerbaijan
##	NA	8.428523	NA
##	Bangladesh	Benin	Bolivia
##	NA	19.200608	2.192031
##	Burkina Faso	Burundi	Cambodia
##	16.263456	NA	NA
##	Cameroon	Chad	Colombia
##	5.020458	NA	NA
##	Comoros	Congo (Brazzaville)	Congo Democratic Republic
##	NA	NA	2.969848
##	Cote d'Ivoire	Dominican Republic	Egypt
##	NA	1.285820	NA
##	Eritrea	Ethiopia	Gabon
##	NA	10.598742	NA
##	Gambia	Ghana	Guinea
##	NA	8.273149	7.283200
##	Guyana	Haiti	Honduras
##	NA	9.124326	1.202082
##	India	Indonesia	Jordan
##	NA	3.265476	NA

##	Kenya	Kyrgyz Republic	Lesotho
##	6.151829	NA	7.212489
##	Liberia	Madagascar	Malawi
##	9.475231	3.394113	5.844941
##	Maldives	Mali	Mauritania
##	NA	10.305824	NA
##	Moldova	Morocco	Mozambique
##	NA	NA	NA
##	Namibia	Nepal	Nicaragua
##	NA	6.018582	NA
##	Niger	Nigeria	Pakistan
##	9.758074	NA	NA
##	Peru	Philippines	Rwanda
##	NA	2.400694	5.729747
##	Sao Tome and Principe	Senegal	Sierra Leone
##	NA	4.605341	2.262742
##	Tajikistan	Tanzania	Timor-Leste
##	NA	NA	NA
##	Togo	Turkmenistan	Uganda
##	NA	NA	NA
##	Ukraine	Vietnam	Yemen
##	NA	NA	NA
##	Zambia	Zimbabwe	
##	25.501046	5.350701	

```
tapply(dhs$beat_goesout, dhs$country, range)
```

```
## $Albania
## [1] 18.6 18.6
##
## $Armenia
## [1] 3.1 19.9
##
## $Azerbaijan
## [1] 42.5 42.5
##
## $Bangladesh
## [1] NA NA
##
## $Benin
## [1] 7.7 44.0
##
## $Bolivia
## [1] 5.9 9.0
##
## $'Burkina Faso'
## [1] 30.3 53.3
##
## $Burundi
## [1] NA NA
##
## $Cambodia
## [1] NA NA
##
```

```

## $Cameroon
## [1] 26.5 33.6
##
## $Chad
## [1] NA NA
##
## $Colombia
## [1] NA NA
##
## $Comoros
## [1] 28.4 28.4
##
## $'Congo (Brazzaville)'
## [1] NA NA
##
## $'Congo Democratic Republic'
## [1] 49.1 53.3
##
## $'Cote d'Ivoire'
## [1] NA NA
##
## $'Dominican Republic'
## [1] 0.8 3.2
##
## $Egypt
## [1] NA NA
##
## $Eritrea
## [1] 51.7 51.7
##
## $Ethiopia
## [1] 43.2 64.2
##
## $Gabon
## [1] NA NA
##
## $Gambia
## [1] 42 42
##
## $Ghana
## [1] 22.3 34.0
##
## $Guinea
## [1] 72.4 82.7
##
## $Guyana
## [1] NA NA
##
## $Haiti
## [1] 10.8 28.8
##
## $Honduras
## [1] 4.3 6.0
##

```

```

## $India
## [1] 29 29
##
## $Indonesia
## [1] 18.2 24.0
##
## $Jordan
## [1] NA NA
##
## $Kenya
## [1] 30.7 39.4
##
## $'Kyrgyz Republic'
## [1] 23.1 23.1
##
## $Lesotho
## [1] 14.2 24.4
##
## $Liberia
## [1] 28.5 41.9
##
## $Madagascar
## [1] 14.4 19.2
##
## $Malawi
## [1] 5.4 16.6
##
## $Maldives
## [1] 13 13
##
## $Mali
## [1] 55.0 74.9
##
## $Mauritania
## [1] NA NA
##
## $Moldova
## [1] 7.3 7.3
##
## $Morocco
## [1] 49.5 49.5
##
## $Mozambique
## [1] NA NA
##
## $Namibia
## [1] NA NA
##
## $Nepal
## [1] 0.5 12.2
##
## $Nicaragua
## [1] 5.8 5.8
##

```



```

## $Niger
## [1] 42.7 56.5
##
## $Nigeria
## [1] NA NA
##
## $Pakistan
## [1] NA NA
##
## $Peru
## [1] NA NA
##
## $Philippines
## [1] 4.3 8.8
##
## $Rwanda
## [1] 26.3 36.6
##
## $'Sao Tome and Principe'
## [1] 10.2 10.2
##
## $Senegal
## [1] 39.9 49.7
##
## $'Sierra Leone'
## [1] 49.7 52.9
##
## $Tajikistan
## [1] 50.5 50.5
##
## $Tanzania
## [1] NA NA
##
## $'Timor-Leste'
## [1] 72.2 72.2
##
## $Togo
## [1] 17.8 17.8
##
## $Turkmenistan
## [1] 40.3 40.3
##
## $Uganda
## [1] NA NA
##
## $Ukraine
## [1] 0.4 0.4
##
## $Vietnam
## [1] NA NA
##
## $Yemen
## [1] 35.5 35.5
##

```

```
## $Zambia
## [1] 29.7 78.8
##
## $Zimbabwe
## [1] 22.3 33.0
```

There is a lot more missing data for these analyses and results appear to vary greatly.

Partie D: Statistiques bivariées

Question 1

Y'a-t-il une association entre la variable `beat_goesout` et `sec_school`? quel type d'analyse vous permet de répondre à cette question?

```
#ctable(dhs$beat_goesout, dhs$sec_school, "r")

# normalement je fais une regression bivariate
```

Question 2

Y'a-t-il une association entre la variable `beat_goesout` et `region`? quel type d'analyse vous permet de répondre à cette question?

```
#ctable(dhs$beat_goesout, dhs$region, "r")

# normalement je fais une regression bivariate, mais si, il y a une association
```

Partie E: Création de votre propre base de données

Question 1

Vous vous demandez d'où provient cette base de données. Et bien comme je vous l'ai dit, cela provient des enquêtes démographiques et de santé. Utiliser Statcompiler pour créer votre base de données sur le sujet qui vous intéresse. Le site est <https://www.statcompiler.com/en/>.

1. Sélectionner l'ensemble des pays
2. Choisissez les deux variables suivantes:

- **Physical or sexual violence committed by husband/partner**
- **Women with secondary or higher education**
- Quelle est la variable dépendante? Quelle est la variable indépendante?

3. Exportez les données dans une base de données (voir Export en haut du site, choisissez database).
4. Ouvrez cette base de données dans R (Quelle est le problème avec votre procédure?)
5. Sélectionner les données du Kenya.

Votre réponse

```
# Malawi

# IV = Women with secondary or higher education
# DV = Physical or sexual violence committed by husband/partner

malawi <- read.csv("../Données/STATcompilerExport202168_19457.xlsx")

## Warning in read.table(file = file, header = header, sep = sep, quote = quote, :
## line 2 appears to contain embedded nulls

## Warning in read.table(file = file, header = header, sep = sep, quote = quote, :
## line 4 appears to contain embedded nulls

## Warning in scan(file = file, what = what, sep = sep, quote = quote, dec = dec, :
## embedded nul(s) found in input

# Il y a les variables nulls

kenya <- read.csv("../Données/STATcompilerExport202168_194837.xlsx", fileEncoding = "UTF-8")

## Warning in read.table(file = file, header = header, sep = sep, quote
## = quote, : invalid input found on input connection '../Données/
## STATcompilerExport202168_194837.xlsx'

## Warning in read.table(file = file, header = header, sep = sep, quote = quote, :
## line 2 appears to contain embedded nulls

## Warning in read.table(file = file, header = header, sep = sep, quote =
## quote, : incomplete final line found by readTableHeader on '../Données/
## STATcompilerExport202168_194837.xlsx'
```

BON EXERCICE
