

# Formation et développement IPUMS: la demande de données

## (Exercice 2 pour R)



### Objectifs

- Créer et télécharger un extrait de données IPUMS PMA
- Décompresser le fichier de données et lire les données dans Stata
- Analyser les données en utilisant un exemple de code

### Résumé

Comprendre la structure des ensembles de données de points de livraison de services IPUMS PMA et la manière dont ils peuvent être exploités pour explorer vos intérêts de recherche. Cet exercice utilisera ensemble de données IPUMS PMA pour explorer la relation entre les caractéristiques des points de livraison de services et en rupture de stock de contraceptifs. Vous allez créer un extrait de données avec les variables EAID, FACILITYTYPEGEN, FACILITYADV, PILLOSBS, PILLOUTDAY, et URBAN; puis vous allez utiliser le code d'exemple à analyser les données.

## Code R à réviser

Les exemples de code et ses réponses dans ce tutoriel utilisent l'extension « tidyverse », mais il y a plusieurs manières à les compléter. Si vous préférez autres styles de programmation, n'hésitez pas à les utiliser. Pour votre référence, voici les explications rapides qui seront utilisés par le tutoriel.

Code	Résultat
%>%	L'opérateur arrange le code de façon emboîtée, ce que facilite la lecture des groupes. Cet opérateur doit être interprété comme « et puis ». L'opérateur organise le code de la suite: ingrédients % > % mélanger () %>% cuire () est équivalent à cuire (mélanger (ingrédients)) (ainsi que "prendre les ingrédients et puis les mélanger et puis les faire cuire").
summarize	Résumer les observations parmi l'ensemble des données dans un ou plusieurs groupes
group_by	Régler les groupes à combiner avec la fonction « summarize »
filter	sélectionne des données d'une table selon une condition.
mutate	Ajouter un nouveau variable à une table
weighted.mean	Calculer la moyenne pondérée
count	La fréquence d'un variable ou valeur
ggplot	Initialiser une visualisation (histogramme, diagramme en boîte à moustaches, etc.)

## Erreurs courantes à éviter

- Ne pas changer le répertoire de travail ou vos données sont stockées
- Confondre = avec pour attribuer une valeur lors de la génération d'une variable, utilisez « <- » (ou « == »). Utilisez « == » comme test logique de l'égalité des objets.



# Enregistrement avec IPUMS

- Allez sur <http://pma.ipums.org>, cliquez sur « Register » pour utiliser IPUMS PMA sur le côté gauche de l'écran. Cliquez sur le bouton « Register for IPUMS PMA » et remplissez le formulaire pour demander l'accès. Vous devrez attendre que votre compte soit approuvé pour accéder aux données. Une fois que vous avez reçu l'email d'approbation, cliquez sur « Connect » en haut de la page et utilisez votre email et votre mot de passe.
- Aller à « Browse and Select Data »
- Choisissez l'unité d'analyse « Person »

CHOOSE THE UNIT OF ANALYSIS FOR DATA BROWSING	
PERSON	EACH RECORD WILL BE A PERSON  DESCRIPTION
SERVICE DELIVERY POINT	EACH RECORD WILL BE A SERVICE DELIVERY POINT  DESCRIPTION

- Cliquez sur la case « Select Samples », cochez la case pour Kenya 2016 R5.

Kenya       2016 R5       2015b R4       2014b R2  
 2015a R3       2014a R1

- Faites défiler vers le bas de la page et cliquez sur l'option du bouton radio pour « All Cases (Respondents and Non-respondents to Service Delivery Point Questionnaires) ». La valeur par défaut est Facility Respondents.
- Cliquez sur la case « Submit Sample Selections. »

## Sample Members

Facility Respondents  
 All Cases (Respondents and Non-respondents to Service Delivery Point Questionnaires)

À l'aide du menu déroulant ou de la fonction de recherche, sélectionnez les variables suivantes:

EAID: Zone de dénombrement (unité d'échantillonnage primaire)

FACILITYTYPEGEN: Type d'établissement



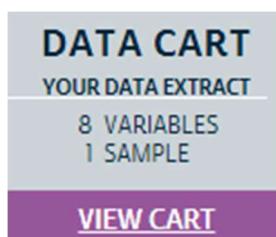
FACILITYADV: Etablissement avancé

PILULES: Pilules contraceptives observés et en stock ou en rupture de stock

PILLOUTDAY: Nombre de jours que les pilules contraceptives ont été en rupture de stock

URBAN: Statut urbain ou rural

- Cliquez sur le bouton violet VIEW CART sous votre panier de données.



- Vérifier la sélection des variables. Notez que certaines variables apparaissent dans votre panier même si vous ne les avez sélectionnés pas. Ils ne sont pas inclus dans le compte de variables dans votre panier de données. Ces variables présélectionnées sont nécessaires pour la pondération, l'estimation de la variance, ou pour identifier l'année, le pays, et le round d'un échantillon.
- Cliquez sur le bouton « Create Data Extract ».
- A côté de « Data Format », cliquez sur « Change ».
- Sélectionnez Stata (.dta), puis cliquez sur Submit»
- Révisez l'écran « Extract Request Summary », décrivez votre extrait et cliquez sur « Submit Extract »
- Vous recevrez un email lorsque les données seront disponibles pour le téléchargement.
- Pour accéder à la page pour télécharger les données, suivez le lien dans l'e-mail ou le lien « My Data Extracts » sur la page d'accueil.

## Lancer les données dans votre logiciel de statistiques

- Allez sur <http://pma.ipums.org> et cliquez sur « My Data Extracts »

Extract Number	Date	Formatted Data	Fixed-width Text Files				Codebook <small>i</small>		
			Data	Command Files <small>i</small>	R	Basic			
51	2018-10-26	--	<a href="#">Download DAT</a>	<a href="#">SPS</a>	<a href="#">SAS</a>	<a href="#">STATA</a>	<a href="#">R</a>	<a href="#">Basic</a>	<a href="#">DDI</a>



- Faites un clic droit sur le lien de données à côté de l'extrait que vous avez créé.
- Choisissez « Save Target As ... » (ou « Save Link As ... »)
- Enregistrer dans « Documents » (qui devrait apparaître comme localisation par défaut)
- Faites la même opération pour le lien DDI à côté de l'extrait
- (Option) Faites la même opération pour le script R
- Ce n'est pas nécessaire décompresser les données pour les utiliser dans R.

Installer et lancer modules R

- Ouvrir R à partir du menu de démarrage.
- Si vous ne l'avez pas encore installé les suites, saisissez:

```
install.packages("ipumsr")
install.packages("dplyr")
install.packages("ggplot2")
```

- Ensuite (ou si vous avez déjà installé les modules sur votre ordinateur), saisissez:

```
library(ipumsr)
library(dplyr)
library(ggplot2)
options(tibble.print_max = Inf)
```

## Lire des données dans R

- Définissez le répertoire de travail où vous avez sauvegardé les données en adaptant la commande suivante (Utilisateurs de Rstudio peuvent aussi utiliser la fonctionnalité « Project » afin d'établir le répertoire de travail. Dans la barre de menu, sélectionnez « File -> New Project -> Existing Directory » et puis parcourez au fichier):

```
setwd("~/") # "~/ " correspond au répertoire « Documents » sur
la plupart des ordinateurs
```



- Exécuter la commande suivante de la console, en adaptant à l'extrait que vous venez de créer (remplacer les #s en-dessus avec les chiffres de votre extrait):

```
ddi <- read_ipums_ddi("pma_000##.xml")
HHF <- read_ipums_micro(ddi)
```

- NOTE: Par cohérence avec les exercices des autres progiciels de statistiques, cet exercice ne passe beaucoup de temps sur la traduction des étiquettes de valeur IPUMS a la base du code R. Veuillez rechercher la commande « vignette » pour se renseigner sur les étiquettes de valeur. Dans R exécutez la commande suivante: `vignette ("value-labels", package = "ipumsr")`



# Analyser l'échantillon

## Partie 1: Explorer les types d'établissements de santé

1. Créez un tableau de fréquences pour FACILITYTYPEGEN indiquant la proportion de chaque type d'établissement enquêtée au Kenya, série 5.

```
PLS %>%
  count(type <- as_factor(FACILITYTYPEGEN)) %>%
  mutate(prop = prop.table(n))
```

2. Selon l'onglet « Universe », quels établissements sont inclus dans l'univers étudié pour FACILITYTYPEGEN? \_\_\_\_\_
3. Les utilisateurs doivent noter que de nombreuses variables de l'enquête sur les points de livraison de services (PLS) ont un univers défini par FACILITYADV, une désignation spécifique au pays de type « d'établissements avancés ». Créez un tableau croisé pour voir quels types d'établissements de la question précédente ont été désignés « établissements avancés » au Kenya pour 2016.

```
PLS %>%
  mutate(ADVANCED = FACILITYADV==1) %>%
  group_by(as_factor(FACILITYTYPEGEN), ADVANCED) %>%
  summarize()
```

4. Consultez l'onglet « Comparability » pour FACILITYADV, en prenant soin de noter que les désignations d'établissements avancés varient d'un pays à l'autre et varient parfois d'une enquête à l'autre dans un pays. Recherchez l'entrée Kenya et déterminez si sa désignation d'établissement avancé correspond à ce que vous avez trouvé à la question 3. La désignation est-elle concordante pour toutes les enquêtes Kényanes comprenant cette variable ? \_\_\_\_\_



## Partie 2: Statistiques descriptives

5. Considérons la variable PILLOBS, qui indique si le PLS disposait d'un stock de pilules contraceptives observable le jour de l'entretien. Selon l'onglet « Codes », quelles sont les réponses possibles pour les PLS enquêtés au Kenya 2016? \_\_\_\_\_
- 
6. Selon l'onglet « Comparability », les réponses possibles à PILLOBS peuvent varier d'un échantillon à l'autre. Comment? \_\_\_\_\_
7. Selon l'onglet « Universe », quels établissements sont inclus dans l'univers de l'enquête pour PILLOBS? \_\_\_\_\_
- 
8. Parmi les établissements qui fournissent généralement les pilules contraceptives indiquées dans PILLOBS, quel type d'établissement était le moins susceptible d'avoir des réserves de pilules contraceptives en stock le jour de l'entretien ? Quelle proportion d'établissements de ce type était en rupture de stock ? (Limitez l'analyse aux seules interviews terminées et aux cas dans l'univers). \_\_\_\_\_
- 

PLS%>%

```
filter(PILLOBS < 90)%>%  
count(FACILITYTYPEGEN, PILLOBS)%>%  
group_by(FACILITYTYPEGEN)%>%  
mutate(type = as_factor(FACILITYTYPEGEN))%>%  
mutate(obs = as_factor(PILLOBS))%>%  
mutate(prop_type = prop.table(n))%>%  
select(type, obs, n, prop_type)
```



## Partie 3: Visualisation des données

9. Pour les établissements qui étaient en rupture de stock de pilules contraceptives, PILLOUTDAY indique le nombre de jours pendant lesquels les réserves étaient indisponibles. Pour certains PLS étant en rupture de stock depuis plus de 90 jours, les codes NIU et les valeurs manquantes pour PILLOUTDAY sont codés sous les valeurs 9994, 9997 et 9999 afin de dépasser la gamme de réponses valides.

Calculez la pénurie moyenne de pilules pour tous les établissements dans l'univers de PILLOUTDAY, en prenant soin d'exclure toute valeur supérieure à 9 000. Recherchez ensuite la moyenne de chaque type d'établissement dans FACILITYTYPEGEN et affichez le résultat sous forme d'histogramme (limitez l'analyse aux réponses valides des PLS dans l'univers pour PILLOUTDAY).

```
PLS%>%
  filter(PILLOUTDAY < 9000)%>%
  summarise(mean(PILLOUTDAY))

PLS%>%
  filter(PILLOUTDAY < 9000)%>%
  group_by(as_factor(FACILITYTYPEGEN))%>%
  summarise(mean(PILLOUTDAY))

PLS%>%
  filter(PILLOUTDAY < 9000)%>%
  group_by(facility_type = as_factor(FACILITYTYPEGEN))%>%
  summarise(mean_days = mean(PILLOUTDAY))%>%
  ggplot() + geom_col(aes(x = facility_type, y = mean_days)) +
  coord_flip()
```



10. Supposons que vous soupçonnez que la différence apparente entre les établissements dans 9 est en réalité une disparité entre les types d'établissements les plus susceptibles de se trouver dans les zones urbaines par rapport aux zones rurales. Créez une paire de graphiques à barres groupé par URBAN pour vérifier si cela est vrai. Existe-t-il des différences entre les établissements urbaines et rurales de chaque type? \_\_\_\_\_

---

```
PLS%>%
  filter(PILLOUTDAY < 9000)%>%
  group_by(facility_type = as_factor(FACILITYTYPEGEN), urban =
as_factor(URBAN))%>%
  summarize(mean_days = mean(PILLOUTDAY))%>%
  ggplot(aes(x = facility_type, y=mean_days)) +
  geom_col(aes(fill = urban), position = position_dodge()) +
  coord_flip()
```

## Partie 4: Combinaison de données PLS et HHF

Les utilisateurs doivent noter que PMA2020 a enquêté des établissements situés dans les mêmes zones d'échantillonnage que les ménages et les femmes lors de la même enquête. Ces données PLS ne sont pas censées être représentatives au niveau national. Au lieu de cela, ils sont destinés à décrire l'environnement de prestation de soins de santé des ménages et des femmes interrogées. Ainsi, il n'y a pas de poids d'échantillonnage pour les variables PLS.

Les fichiers contiennent un poids pour les unités d'échantillonnage EAWEIGHT, qui est un poids de probabilité représentant la possibilité que le secteur de dénombrement (EA) soit sélectionné pour l'échantillonnage. Les collecteurs des données originales ne recommandent pas d'utiliser EAWEIGHT pour pondérer les variables PLS. La meilleure



utilisation des variables PLS consiste plutôt de calculer des statistiques récapitulatives au niveau des SD et à les associer au jeu de données Ménage and Femme (HHF) à l'aide de la variable EAID en tant que source d'informations contextuelles pour l'environnement de prestation de services de chaque femme.

Par exemple, vous pouvez utiliser les variables PILLOBS et PILLOUTDAY pour déterminer si un établissement dans chaque EAID est en rupture de stock de pilules contraceptives et le nombre moyen de jours pendant lesquels les établissements de chaque EAID ont été en rupture de stock. Ces statistiques récapitulatives peuvent être combinées avec l'ensemble de données HHF afin de montrer si chaque femme répondante avait un accès local fiable aux pilules contraceptives.

11. Générez un tableau avec le nombre des femmes âgées de 15 à 49 ans (ÉLIGIBLE == 1) de l'échantillon Kenya 2016 de la série 5 HHF qui résidaient dans chacun des secteurs de dénombrement où les pilules contraceptives n'étaient pas disponibles dans tous les établissements locaux de l'enquête PLS? Combien de secteurs de dénombrement de Kenya 2016 satisfont à ces critères? \_\_\_\_\_

---

```
HHF%>%
```

```
  mutate(pillobs = HHF$EAID %in% subset(PLS$EAID, PLS$PILLOBS ==  
  3)) %>%  
  
  group_by(EAID) %>%  
  
  filter(pillobs==TRUE & ELIGIBLE==1) %>%  
  
  count(pillobs) %>%  
  
  select(EAID, n)
```

12. A l'aide du tableau de la question dernière, quel est le nombre total des femmes âgées de 15 à 49 ans (ÉLIGIBLE == 1) de l'échantillon Kenya 2016 de la série 5 de l'ensemble de données Ménage et Femme (HHF) qui résidaient dans chacun des



secteurs de dénombrement où les pilules contraceptives n'étaient pas disponibles dans tous les établissements locaux de l'enquête PLS?

---

13. Effectuez une régression logistique pour estimer le lien entre les femmes qui prennent actuellement la pilule (FPNOWUSPILL) et la durée moyenne de la pénurie (PILLOUTDAY) pour chaque secteur de dénombrement qui était en absence de pilules le jour de l'interview PLS. Ajustez votre modèle pour qu'il soit représentatif des femmes kényanes utilisant FQWEIGHT. Considérez l'effet du nombre moyen de jours en rupture de stock comme une variable continue et un prédicteur de l'effet recherché. Recodez les valeurs pour FPNOWUSPILL qui ne sont pas dans l'univers ou qui manquent à zéro.

Existe-t-il un lien entre le nombre de jours pendant lesquels les établissements du secteur de dénombrement de la femme sont en rupture de stock de pilules et l'utilisation actuelle de la pilule par la femme pour la planification familiale?

```
model_data <- HHF%>%
  left_join(PLS%>%
    group_by(EAID)%>%
    summarize(pilloutday = mean(subset(PILLOUTDAY,
PILLOUTDAY < 9000))))%>%
  mutate(fpnowuspill = case_when(as.numeric(FPNOWUSPILL) > 90 ~
0, TRUE ~ as.numeric(FPNOWUSPILL)))%>%
  mutate(pilloutday = case_when(is.na(pilloutday) ~ 0, TRUE ~
pilloutday))

model <- glm(FPNOWUSPILL ~ pilloutday, data = model_data, family =
binomial, weights = round(FQWEIGHT))
summary(model)

exp(coef(model))
```



# RÉPONSES

## Partie 1: Explorer les types d'établissements de santé

- Créez un tableau de fréquences pour FACILITYTYPEGEN indiquant la proportion de chaque type d'établissement enquêtée au Kenya, série 5.

```
> SDP %>%
+   count(type <- as_factor(FACILITYTYPEGEN)) %>%
+   mutate(prop = prop.table(n))
# A tibble: 7 x 3
`type <- as_factor(FACILITYTYPEGEN)`      n     prop
<fct>                               <int>    <dbl>
1 Hospital                           79 0.185
2 Health center                      90 0.210
3 Health clinic                      16 0.0374
4 Other health facility              1   0.00234
5 Dispensary                         190 0.444
6 Pharmacy/chemist/drug shop        48 0.112
7 other                             4  0.00935
```

- Selon l'onglet « Universe », quels établissements sont inclus dans l'univers étudié pour FACILITYTYPEGEN? Tous les points de livraison de services
- Les utilisateurs doivent noter que de nombreuses variables de l'enquête sur les points de prestation de services (PLS) ont un univers défini par FACILITYADV, une désignation spécifique au pays de types « établissements avancés ». Créez un tableau croisé pour voir quels types d'établissements de la question précédente ont été désignées comme « établissements avancés » au Kenya pour 2016. Toutes sont avancées, à l'exception des pharmacies.

```
> SDP %>%
+   mutate(ADVANCED = FACILITYADV==1)%>%
+   group_by(as_factor(FACILITYTYPEGEN), ADVANCED)%>%
+   summarize()
# A tibble: 7 x 2
# Groups:   as_factor(FACILITYTYPEGEN) [?]
`as_factor(FACILITYTYPEGEN)` ADVANCED
<fct>                  <lgl>
1 Hospital                TRUE
2 Health center            TRUE
3 Health clinic            TRUE
4 Other health facility   TRUE
5 Dispensary               TRUE
6 Pharmacy/chemist/drug shop FALSE
7 other                   TRUE
```



4. Consultez l'onglet « Comparability » pour FACILITYADV, en prenant soin de noter que les désignations d'établissements avancés varient d'un pays à l'autre et varient parfois d'une enquête à l'autre dans un pays. Recherchez l'entrée Kenya et déterminez si sa désignation d'établissement avancé correspond à ce que vous avez trouvé à la question 3. La désignation est-elle concordante pour toutes les enquêtes Kényanes comprenant cette variable ? Ils correspondent, et toutes les enquêtes kényanes interrogés ont la même désignation.

## Partie 2: Statistiques descriptives

5. Considérons la variable PILLOBS, qui indique si le PLS disposait d'un stock de pilules contraceptives observable le jour de l'entretien. Selon l'onglet « Codes », quelles sont les réponses possibles pour les PLS enquêtés au Kenya 2016?

<u>1 - In-stock and observed</u>	<u>94 - Not interviewed (PLS questionnaire)</u>
<u>2 - In-stock but not observed</u>	<u>98 - No response or missing</u>
<u>3 - Out of stock</u>	<u>99 - NIU (not in universe)</u>

6. Selon l'onglet « Comparability », les réponses possibles à PILLOBS peuvent varier d'un échantillon à l'autre. Comment? Certains échantillons initiaux incluent moins de détails et fournissent des réponses dichotomiques selon que l'enquêteur a observé la pilule contraceptive en stock. Dans ces échantillons initiaux, si les pilules contraceptives n'étaient pas observées, elles étaient supposées être "en rupture de stock". Lors d'enquêtes ultérieures, les enquêteurs avaient l'option de signaler que les pilules contraceptives étaient "en stock mais pas observées".

7. Selon l'onglet « Universe », quels établissements sont inclus dans l'univers de l'enquête pour PILLOBS? Services de livraison fournissant des pilules contraceptives.



8. Parmi les établissements qui fournissent généralement les pilules contraceptives indiquées dans PILLOBS, quel type d'établissement était le moins susceptible d'avoir des réserves de pilules contraceptives en stock le jour de l'entretien? Quelle proportion d'établissements de ce type était en rupture de stock? (Limitez l'analyse aux seules interviews terminées et aux cas dans l'univers). Les cliniques de santé étaient les plus susceptibles de manquer de médicaments avec 25% de rupture de stock.

```
> SDP%>%
+   filter(PILLOBS < 90)%>%
+   count(FACILITYTYPEGEN, PILLOBS)%>%
+   group_by(FACILITYTYPEGEN)%>%
+   mutate(type = as_factor(FACILITYTYPEGEN))%>%
+   mutate(obs = as_factor(PILLOBS))%>%
+   mutate(prop_type = prop.table(n))%>%
+   select(type, obs, n, prop_type)
Adding missing grouping variables: `FACILITYTYPEGEN`
# A tibble: 17 x 5
# Groups:   FACILITYTYPEGEN [7]
  FACILITYTYPEGEN type      obs        n prop_type
  <int+lbl>    <fct>    <fct>    <int>    <dbl>
1 1             Hospital In-stock and observed 69  0.896
2 1             Hospital In-stock but not observed 1  0.0130
3 1             Hospital Out of stock          7  0.0909
4 2             Health center In-stock and observed 69  0.812
5 2             Health center In-stock but not observed 2  0.0235
6 2             Health center Out of stock          14 0.165
7 3             Health clinic In-stock and observed 8  0.667
8 3             Health clinic In-stock but not observed 1  0.0833
9 3             Health clinic Out of stock          3  0.25
10 4            Other health facility In-stock and observed 1  1
11 6            Dispensary In-stock and observed 144 0.783
12 6            Dispensary In-stock but not observed 3  0.0163
13 6            Dispensary Out of stock          37 0.201
14 7            Pharmacy/chemist/drug shop In-stock and observed 36 0.878
15 7            Pharmacy/chemist/drug shop In-stock but not observed 2  0.0488
16 7            Pharmacy/chemist/drug shop Out of stock          3  0.0732
17 9            other     In-stock and observed 1  1
```

### Partie 3: Visualisation des données

9. Pour les établissements qui... Calculez la pénurie moyenne de pilules pour tous les établissements dans l'univers de PILLOUTDAY, en prenant soin d'exclure toute valeur supérieure à 9 000. Recherchez ensuite la moyenne de chaque type d'établissement dans FACILITYTYPEGEN et affichez le résultat sous forme d'histogramme (limitez l'analyse aux réponses valides des PLS dans l'univers pour PILLOUTDAY).

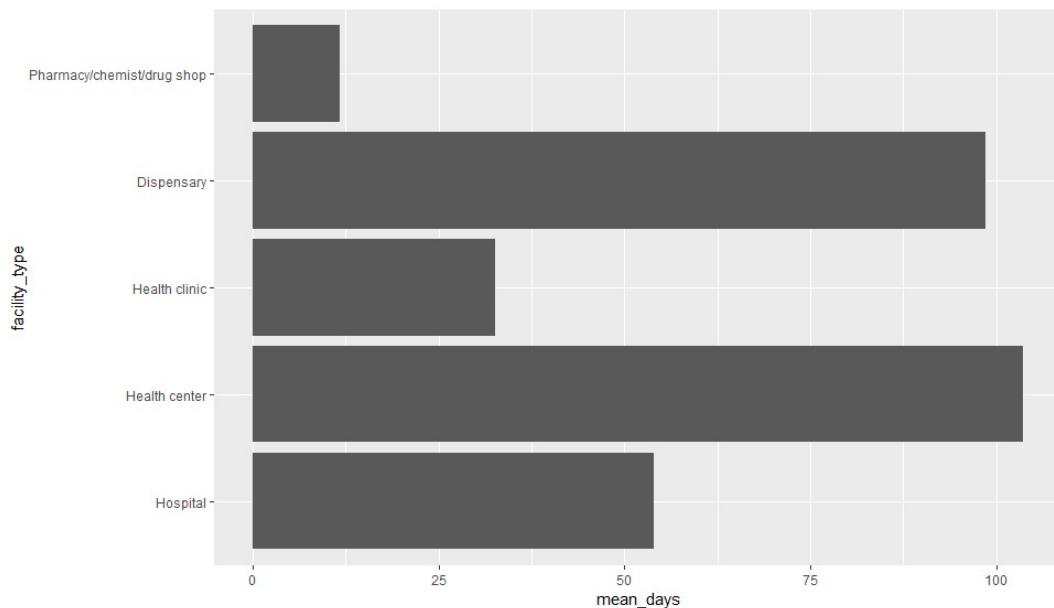
```
> SDP%>%
+   filter(PILLOUTDAY < 9000)%>%
+   summarise(mean(PILLOUTDAY))
# A tibble: 1 x 1
`mean(PILLOUTDAY)` <dbl>
1                  87.5
```



```

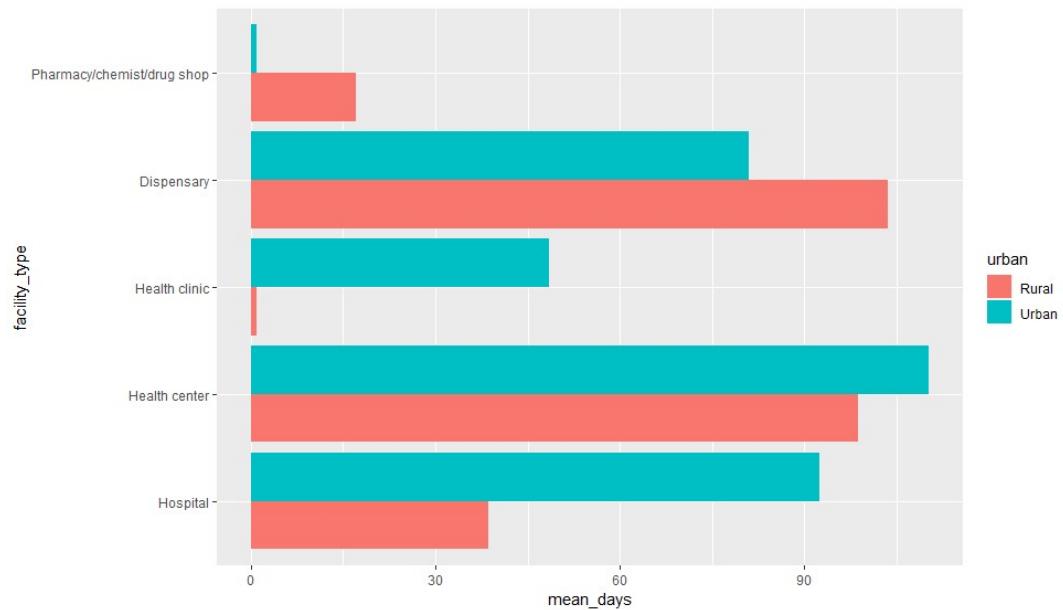
> SDP%>%
+   filter(PILLOUTDAY < 9000)%>%
+   group_by(facility_type = as_factor(FACILITYTYPEGEN), urban = as_factor(URBAN))%>%
+   summarize(mean(PILLOUTDAY))
# A tibble: 10 x 3
# Groups: facility_type [?]
  facility_type    urban `mean(PILLOUTDAY)`
  <fct>        <fct>            <dbl>
1 Hospital       Rural      38.6
2 Hospital       Urban     92.5
3 Health center  Rural     98.8
4 Health center  Urban    110.
5 Health clinic  Rural      1
6 Health clinic  Urban    48.5
7 Dispensary     Rural    104.
8 Dispensary     Urban    80.9
9 Pharmacy/chemist/drug shop Rural     17
10 Pharmacy/chemist/drug shop Urban      1

```



10. Supposons que vous soupçonnez que la différence apparente entre les établissements d'A est en réalité une disparité entre les types d'établissements les plus susceptibles de se trouver dans les zones urbaines par rapport aux zones rurales. Créez une paire de graphiques à barres groupé par URBAN pour vérifier si cela est vrai. Y-t-il des différences entre les établissements urbaines et rurales de chaque type?





## Partie 4: Combinaison de données PLS et HHF

11. Générez un tableau avec le nombre des femmes âgées de 15 à 49 ans (ÉLIGIBLE == 1) de l'échantillon Kenya 2016 de la série 5 HHF qui résidaient dans chacun des secteurs de dénombrement où les pilules contraceptives n'étaient pas disponibles dans tous les établissements locaux de l'enquête PLS? Combien de secteurs de dénombrement de Kenya 2016 satisfont à ces critères?

43 secteurs de dénombrement



```

> HHF%>%
+   mutate(pillobs = HHF$EAID %in% subset(SDP$EAID, SDP$PILLOBS == 3))%>%
+   group_by(EAID)%>%
+   filter(pillobs==TRUE & ELIGIBLE==1)%>%
+   count(pillobs)%>%
+   select(EAID, n)
# A tibble: 43 x 2
# Groups:   EAID [43]
  EAID     n
<dbl> <int>
1 4013    33
2 4047    44
3 4163    47
4 4207    38
5 4212    23
6 4214    38
7 4234    50
8 4245    45
9 4318    36
10 4336   41
11 4356    47
12 4361    45
13 4373    39
14 4431    40
15 4456    50
16 4485    51
17 4531    42
18 4626    23
19 4628    27
20 4639    36
21 4655    39
22 4662    34
23 4676    56
24 4677    44
25 4707    56
26 4711    44
27 4712    22
28 4719    46
29 4760    52
30 4768    29
31 4784    50
32 4795    63
33 4871    51
34 4885    25
35 4887    45
36 4895    55
37 4899    34
38 4905    38
39 4938    36
40 4952    54
41 4963    39
42 4965    36
43 4974    34

```

12. A l'aide du tableau de la question dernière, quel est le nombre total des femmes âgées de 15 à 49 ans (`ÉLIGIBLE == 1`) de l'échantillon Kenya 2016 de la série 5 de l'ensemble de données Ménage et Femme (HHF) qui résidaient dans chacun des secteurs de dénombrement où les pilules contraceptives n'étaient pas disponibles dans tous les établissements locaux de l'enquête PLS?

Ce sont les trois premiers:

EAID 4013 = 33 femmes

EAID 4163 = 47 femmes

EAID 4047 = 44 femmes



13. Effectuez une régression logistique pour estimer ... Existe-t-il un lien entre le nombre de jours pendant lesquels les établissements du secteur de dénombrement de la femme sont en rupture de stock de pilules et l'utilisation actuelle de la pilule par la femme pour la planification familiale?

La probabilité qu'une femme de l'échantillon utilise des pilules contraceptives reste la même quel que soit le nombre moyen de jours pendant lesquels son PLS local n'a pas été disponible (rapport de cotes = 1,000). Cependant, ce résultat n'est pas statistiquement significatif ( $p = 0,728$ , IC à 95% entre 0,998 et 1,002).

FPNOWUSPILL	Rapport de cotes	Valeur p	IC à 95%
PILLOUTDAY	1,000	0,728	(0,998 et 1,002)

```
> model_data <- HHF%>%
+   left_join(SDP)%>%
+     group_by(EAID)%>%
+       summarize(pilloutday = mean(subset(PILLOUTDAY, PILLOUTDAY < 9000))))%>%
+   mutate(fpnowuspill = case_when(as.numeric(FPNOWUSPILL) > 90 ~ 0, TRUE ~ as.numeric(FPNOWUSPILL)))%>%
+   mutate(pilloutday = case_when(is.na(pilloutday) ~ 0, TRUE ~ pilloutday))
Joining, by = "EAID"
>
> model <- glm(FPNOWUSPILL ~ pilloutday, data = model_data, family = binomial, weights = round(FQWEIGHT))
> summary(model)

call:
glm(formula = FPNOWUSPILL ~ pilloutday, family = binomial, data = model_data,
    weights = round(FQWEIGHT))

Deviance Residuals:
    Min      1Q  Median      3Q     Max 
-0.5405  0.0000  0.0000  0.0000   4.2755 

Coefficients:
            Estimate Std. Error z value Pr(>|z|)    
(Intercept) -2.9979074  0.0649217 -46.177  <2e-16 ***
pilloutday  -0.0003712  0.0010675  -0.348    0.728  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 2378.5  on 5521  degrees of freedom
Residual deviance: 2378.4  on 5520  degrees of freedom
AIC: 2382.4

Number of Fisher Scoring iterations: 5

> exp(coef(model))
(Intercept)  pilloutday
 0.04989136  0.99962886
```

