FACULTÉ DES SCIENCES ECONOMIQUES

UNIVERSITÉ DE RENNES 1

Logiciel R et programmation **Master 1 Statistique & Économétrie**

Ewen Gallic [1]

Date : 14 Octobre 2015

Tous documents autorisés

Durée : 1h

Vous avez reçu une archive nommée `NOM_prenom.zip` [2]. Elle contient :

- `cc1.Rproj` : le fichier de projet ;
- `reponses.Rmd` : le fichier dans lequel vous reporterez les réponses aux questions des exercices.

Avant de commencer à rédiger :

1. renommez le répertoire `NOM_Prenom` pour que le nom corresponde à votre nom et votre prénom ;
2. ouvrez le fichier `cc1.Rproj` dans RStudio ;
3. ouvrez le fichier `reponses.Rmd`, dans lequel vous écrirez vos solutions.

Exercice 1

1. Récupérer, à l'aide de la fonction appropriée, le fichier Excel situé à l'adresse suivante : http://egallic.fr/Enseignement/R/2015/cc1/gdp.xls.

```
lk <- "http://egallic.fr/Enseignement/R/2015/cc1/gdp.xls"
download.file(lk, destfile = "gdp.xls", mode = "wb")
```

2. Ce fichier contient des données trimestrielles provenant d'Eurostat. Il est composé de trois feuilles :

   (a) `gdp` : PIB pour l'europe des 25, l'Allemagne, la France et l'Italie, en prix courants, en millions d'euros,

   (b) `gdp_def` : le déflateur du PIB pour les mêmes régions géographiques, avec 2010 comme base 100,

   (c) `pop` : la population en milliers de personnes d'Allemagne, de France et d'Italie.

   Les valeurs manquantes sont indiquées par le caractère `:` (deux points).

   Importer les données des tableaux de chaque feuille dans `R`, afin d'obtenir trois tableaux de données dont le nom correspondra à celui de la feuille du fichier Excel (`gdp`, `gdp_def` et `pop`) ;

```

```

1. ewen.gallic[at]univ-rennes1.fr
2. Elle est disponible à l'URL suivante : http://egallic.fr/Enseignement/R/2015/cc1/NOM_prenom.zip

```r
# donnees du pib
library(readxl)
gdp <- read_excel("gdp.xls", sheet = "gdp",
                  skip = 10, na = ":")
# donnees du deflateur du pib
gdp_def <- read_excel("gdp.xls", sheet = "gdp_def",
                      skip = 10, na = ":")
# donnees de population
pop <- read_excel("gdp.xls", sheet = "pop",
                  skip = 10, na = ":")
```

3. Renommer la colonne `TIME/GEO` du tableau de données `gdp` en `time` ;

```r
library(dplyr)
gdp <-
  gdp %>%
  rename(time = `TIME/GEO`)
```

4. Vérifier que le tableau `gdp` ne contient aucun doublon, et les retirer le cas échéant ;

```r
any(duplicated(gdp))
```

5. Sur chacun des trois tableaux de données (`gdp`, `gdp_def` et `pop`), appliquer la fonction `gather()` du *package* `tidyr` de manière à avoir trois colonnes, afin d'obtenir :
   — pour le tableau `gdp` : `time` (année et trimestre), `geo` (zone géographique) et `gdp` (valeur du PIB),
   — pour le tableau `gdp_def` : `time` (année et trimestre), `geo` (zone géographique) et `gdp_def` (valeur du déflateur du PIB),
   — pour le tableau `pop` : `time` (année et trimestre), `geo` (zone géographique) et `pop` (population) ;

```r
library(tidyr)
gdp <-
  gdp %>%
  gather(geo, gdp, -time)

gdp_def <-
  gdp_def %>%
  gather(geo, gdp_def, -time)

pop <-
  pop %>%
  gather(geo, pop, -time)
```

6. Fusionner les trois tableaux `gdp`, `gdp_def` et `pop` dans un seul que l'on nommera `df`. La fusion doit se faire par les colonnes `time` et `geo` ;

```
df <-
  gdp %>%
  left_join(gdp_def) %>%
  left_join(pop)
```

7. Filtrer le tableau `df` afin de retirer les observations pour lesquelles la variable `geo` prend la valeur `eu` ;

```
df <-
  df %>%
  filter(geo != "eu")
```

8. Ajouter au tableau `df` les variables suivantes :
   — `annee` : l'année de l'observation, avec quatre chiffres, en `numeric`,
   — `trim` : le trimestre de l'observation, en `numeric`,
   — `date` : l'année à laquelle on ajoute 0 pour le premier trimestre, 0.25 pour le second trimestre, 0.50 pour le troisième et 0.75 pour le quatrième ;

```
library(stringr)
df <-
  df %>%
  mutate(annee = str_sub(time, 1, 4) %>% as.numeric(),
         trim = str_sub(time, -1) %>% as.numeric(),
         date = annee + trim/4 - 0.25)
```

9. *Note : le tableau `df` ne doit pas être altéré lors de cette question.*
   Filtrer le tableau `df` pour conserver uniquement les observations pour lesquelles la variable `time` prend la valeur `2012Q2`, sélectionner uniquement les variables `geo` et `pop`, puis renommer la variable `pop` du tableau ainsi obtenu en `pop_ref`.
   Stocker le résultat dans un objet que l'on nommera `reference_pop` ;

```
reference_pop <-
  df %>%
  filter(time == "2010Q2") %>%
  select(geo, pop) %>%
  rename(pop_ref = pop)
```

10. À l'aide de la colonne `geo` dans les tableaux `df` et `reference_pop`, effectuer une fusion des tableaux de données `df` et `reference_pop` pour ajouter le contenu de `reference_pop` dans `df` ;

```
df <-
  df %>%
  left_join(reference_pop)
```

11. Maintenant que le tableau `df` contient la valeur de la population pour l'année de référence (2010), ajouter la colonne `pop_index` à `df` en divisant la population (`pop`) par la population de référence (`pop_ref`). Ajouter par la même occasion la colonne `gdp_real`, soit le PIB réel, défini comme le PIB (`gdp`) divisé par son déflateur (`gdp_def`) ;

```
df <-
  df %>%
  mutate(pop_index = pop / pop_ref,
         gdp_real = gdp / gdp_def)
```

12. Ajouter à `df` la colonne `lny` qui doit indiquer la valeur réelle du PIB auquel on a retiré la tendance de la population. Le calcul doit s'effectuer pour chaque zone géographique i et chaque date t comme suit :

$$\ln y_{it} = 100 \times log\left(\frac{\text{GDP}_{it}}{\text{GDP Def}_i} \times \frac{1}{\text{Pop Index}_{it}}\right),$$

avec y la production, GDP le PIB, GDP Def le déflateur du PIB et Pop Index l'indice de la population ;

```
df <-
  df %>%
  mutate(lny = 100 * log(gdp / gdp_def / pop_index))
```

13. Pour chaque pays (Allemagne, France et Italie), et pour chaque trimestre, en utilisant les données contenues dans `df`, calculer la moyenne et l'écart-type du PIB réel (`gdp_real`). Utiliser les fonctions appropriées du *package* `dplyr` pour effectuer les regroupements.

```
df %>%
  group_by(geo, trim) %>%
  summarize(moyenne_pib = mean(gdp_real, na.rm=TRUE),
            sd_pib = sd(gdp_real, na.rm = TRUE))
```

Exercice 2

Créer une fonction que l'on nommera `afficher_date`, qui, lorsqu'on lui donne une chaîne de caractère contenant une date au format `YYYY-MM-DD` affiche dans la console l'année et le mois de la date, au sein d'une phrase.

La phrase affichée doit être comme dans l'exemple qui suit, avec l'année et le mois correspondant à la chaîne de caractère fournie à la fonction.
Avec `2012-12-21`, la fonction doit afficher dans la console :
`"L'année vaut : 2012 et le mois : 12"`

```
library(lubridate)
# Affiche l'année et le mois d'une date dans la console
# x@ : (string) date au format YYYY-MM-DD
afficher_date <-
  function(x){
  d <- ymd(x)
  print(str_c("L'année vaut : ", year(d),
              " et le mois : ", month(d)))
}

afficher_date("2012-12-21")
```