

Самостійна робота 2

Юрій Харченко

11/7/2020

Продовження роботи, розпочатої в Самостійна робота 1

Підключимо потрібні пакети

```
library(WDI)
library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

Встановимо робочі константи

```
start <- 1990L
end <- 2019L

ind_pop_totl <- "SP.POP.TOTL"
ind_gdp_pcap <- "NY.GDP.PCAP.PP.CD"
```

Завантажимо створені в попередній роботі фрейми

```
# Населення за роками

pop_totl <- read.csv(file = "pop_totl.csv")
head(pop_totl)

##   X year      ukr      pol      rus decade
## 1 1 1990 51891400 38110782 147969407 1990
## 2 2 1991 52000500 38246193 148394216 1990
## 3 3 1992 52150400 38363667 148538197 1990
## 4 4 1993 52179200 38461408 148458777 1990
## 5 5 1994 51921400 38542652 148407912 1990
## 6 6 1995 51512800 38594998 148375787 1990

# ВВП на душу населення за роками
```

```
gdp_pcap <- read.csv(file = "gdp_pcap.csv")
head(gdp_pcap)
```

```
##   X year      ukr      pol      rus decade
## 1 1 1990 7305.143 6173.862 8027.846   1990
## 2 2 1991 6902.486 5913.855 7857.906   1990
## 3 3 1992 6356.734 6181.769 6862.463   1990
## 4 4 1993 5578.521 6548.119 6419.516   1990
## 5 5 1994 4412.745 7027.078 5734.414   1990
## 6 6 1995 3987.006 7662.772 5613.281   1990
```

Країни з великим населенням та їх ВВП на душу за 2019 рік

```
all <- read.csv(file = "all.csv")
head(all)
```

```
##   X iso2c      country SP.POP.TOTL year iso3c      region
## 1 1 AR      Argentina  44938712 2019 ARG Latin America & Caribbean
## 2 2 BD      Bangladesh 163046161 2019 BGD South Asia
## 3 3 BR      Brazil    211049527 2019 BRA Latin America & Caribbean
## 4 4 CD Congo, Dem. Rep. 86790567 2019 COD Sub-Saharan Africa
## 5 5 CN      China     1397715000 2019 CHN East Asia & Pacific
## 6 6 CO      Colombia  50339443 2019 COL Latin America & Caribbean
##   capital longitude latitude      income lending
## 1 Buenos Aires -58.4173 -34.61180 Upper middle income IBRD
## 2 Dhaka         90.4113  23.70550 Lower middle income IDA
## 3 Brasilia      -47.9292 -15.78010 Upper middle income IBRD
## 4 Kinshasa      15.3222  -4.32500 Low income IDA
## 5 Beijing       116.2860 40.04950 Upper middle income IBRD
## 6 Bogota        -74.0820  4.60987 Upper middle income IBRD
## NY.GDP.PCAP.PP.CD
## 1 22947.140
## 2 4950.744
## 3 15258.851
## 4 1143.453
## 5 16784.659
## 6 15643.731
```

Коментар до завантажених даних

Дані отримані за допомогою пакету WDI, який дозволяє робити он-лайн запити до баз даних Світового банку

pop_totl - містить часові ряди 1990-2019рр загальної кількості населення для України, Польщі та Росії

gdp_pcap - містить часові ряди ВВП на душу населення 1990-2019рр для України, Польщі та Росії

all - містить дані 35 країн з найбільшим населенням з такими кількісними показниками:

- SP.POP.TOTL - загальна кількість населення
- NY.GDP.PCAP.PP.CD - ВВП на душу населення

та факторними змінними:

- region - регіон
- income - рівень доходу
- lending - група позик

Додамо до фрейму all ще два кількісних показника

Виберемо з фрейму список кодів країн, по яких будемо завантажувати дані

```
countries <- select(all, iso2c)
str(countries)
```

```
## 'data.frame': 35 obs. of 1 variable:
## $ iso2c: chr "AR" "BD" "BR" "CD" ...
```

```
y <- 2016
```

Загальна площа країни (кв.км)

```
ind_lnd_totl <- "AG.LND.TOTL.K2"
```

```
land_all <- WDI(country=c(countries$iso2c), indicator=ind_lnd_totl, start=y, end=y, extra=TRUE, cache=NU
  arrange(iso2c)
```

```
head(land_all)
```

```
##   iso2c      country AG.LND.TOTL.K2 year iso3c      region
## 1  AR      Argentina    2736690 2016  ARG Latin America & Caribbean
## 2  BD      Bangladesh    130170 2016  BGD      South Asia
## 3  BR      Brazil      8358140 2016  BRA Latin America & Caribbean
## 4  CD Congo, Dem. Rep.    2267050 2016  COD      Sub-Saharan Africa
## 5  CN      China      9388210 2016  CHN      East Asia & Pacific
## 6  CO      Colombia    1109500 2016  COL Latin America & Caribbean
##      capital longitude latitude      income lending
## 1 Buenos Aires -58.4173 -34.6118 Upper middle income IBRD
## 2      Dhaka    90.4113  23.7055 Lower middle income  IDA
## 3    Brasilia -47.9292 -15.7801 Upper middle income  IBRD
## 4    Kinshasa  15.3222  -4.325   Low income      IDA
## 5    Beijing  116.286  40.0495 Upper middle income  IBRD
## 6    Bogota   -74.082  4.60987 Upper middle income  IBRD
```

```
summary(land_all)
```

```
##      iso2c      country      AG.LND.TOTL.K2      year
## Length:35      Length:35      Min.   : 97489      Min.   :2016
## Class :character Class :character 1st Qu.: 398311 1st Qu.:2016
## Mode :character  Mode :character Median : 828340 Median :2016
##                                     Mean : 2126901 Mean : 2016
##                                     3rd Qu.: 1910855 3rd Qu.:2016
##                                     Max.   :16376870 Max.   :2016
##                                     NA's   :1
##      iso3c      region      capital      longitude
## Length:35      Length:35      Length:35      Length:35
## Class :character Class :character Class :character Class :character
## Mode :character  Mode :character Mode :character Mode :character
##
##
##
##
##      latitude      income      lending
```

```
## Length:35      Length:35      Length:35
## Class :character Class :character Class :character
## Mode :character Mode :character Mode :character
##
##
##
##
```

Площа с.г. (кв.км)

```
ind_lnd_agri <- "AG.LND.AGRI.K2"
```

```
lnd_agri_all <- WDI(country=c(countries$iso2c), indicator=ind_lnd_agri, start=y, end=y, extra=TRUE, cache=TRUE,
  arrange(iso2c))
```

```
head(lnd_agri_all)
```

```
##   iso2c      country AG.LND.AGRI.K2 year iso3c      region
## 1   AR      Argentina  1487000.0 2016  ARG Latin America & Caribbean
## 2   BD      Bangladesh   91942.1 2016  BGD      South Asia
## 3   BR          Brazil  2835460.0 2016  BRA Latin America & Caribbean
## 4   CD Congo, Dem. Rep.  262000.0 2016  COD      Sub-Saharan Africa
## 5   CN          China  5277330.0 2016  CHN      East Asia & Pacific
## 6   CO      Colombia   446656.0 2016  COL Latin America & Caribbean
##      capital longitude latitude      income lending
## 1 Buenos Aires -58.4173 -34.6118 Upper middle income IBRD
## 2      Dhaka    90.4113  23.7055 Lower middle income  IDA
## 3   Brasilia -47.9292 -15.7801 Upper middle income IBRD
## 4   Kinshasa  15.3222  -4.325   Low income      IDA
## 5    Beijing  116.286  40.0495 Upper middle income IBRD
## 6    Bogota   -74.082  4.60987 Upper middle income IBRD
```

```
summary(lnd_agri_all)
```

```
##      iso2c      country      AG.LND.AGRI.K2      year
## Length:35      Length:35      Min.   : 17008      Min.   :2016
## Class :character Class :character 1st Qu.: 155360      1st Qu.:2016
## Mode :character Mode :character Median : 368440      Median :2016
##                                     Mean  : 781561      Mean  :2016
##                                     3rd Qu.: 694931      3rd Qu.:2016
##                                     Max.   :5277330      Max.   :2016
##      iso3c      region      capital      longitude
## Length:35      Length:35      Length:35      Length:35
## Class :character Class :character Class :character Class :character
## Mode :character Mode :character Mode :character Mode :character
##
##
##
##      latitude      income      lending
## Length:35      Length:35      Length:35
## Class :character Class :character Class :character
## Mode :character Mode :character Mode :character
##
##
##
```

Населення с.г. (% до загального)

```
ind_rur_totl <- "SP.RUR.TOTL.ZS"

rur_totl_all <- WDI(country=c(countries$iso2c), indicator=ind_rur_totl, start=end, end=end, extra=TRUE,
  arrange(iso2c))

head(rur_totl_all)
```

```
##   iso2c      country SP.RUR.TOTL.ZS year iso3c      region
## 1   AR      Argentina      8.009 2019   ARG Latin America & Caribbean
## 2   BD      Bangladesh     62.595 2019   BGD      South Asia
## 3   BR      Brazil        13.176 2019   BRA Latin America & Caribbean
## 4   CD Congo, Dem. Rep.    54.954 2019   COD      Sub-Saharan Africa
## 5   CN      China         39.692 2019   CHN      East Asia & Pacific
## 6   CO      Colombia      18.896 2019   COL Latin America & Caribbean
##   capital longitude latitude      income lending
## 1 Buenos Aires -58.4173 -34.6118 Upper middle income IBRD
## 2      Dhaka    90.4113  23.7055 Lower middle income  IDA
## 3   Brasilia  -47.9292 -15.7801 Upper middle income IBRD
## 4   Kinshasa  15.3222  -4.325   Low income  IDA
## 5    Beijing  116.286  40.0495 Upper middle income IBRD
## 6    Bogota   -74.082  4.60987 Upper middle income IBRD
```

```
summary(rur_totl_all)
```

```
##   iso2c      country      SP.RUR.TOTL.ZS      year
## Length:35      Length:35      Min.   : 8.009      Min.   :2019
## Class :character Class :character 1st Qu.:19.495 1st Qu.:2019
## Mode  :character Mode  :character Median :33.144 Median :2019
##                                     Mean  :40.115 Mean  :2019
##                                     3rd Qu.:62.844 3rd Qu.:2019
##                                     Max.   :78.775 Max.   :2019
##   iso3c      region      capital      longitude
## Length:35      Length:35      Length:35      Length:35
## Class :character Class :character Class :character Class :character
## Mode  :character Mode  :character Mode  :character Mode  :character
##
##
##   latitude      income      lending
## Length:35      Length:35      Length:35
## Class :character Class :character Class :character
## Mode  :character Mode  :character Mode  :character
##
##
##
```

Приєднаємо отримані показники до основного фрейму та запишемо в файл

```
all1 <- inner_join(all, select(land_all, iso2c, AG.LND.TOTL.K2), by = c("iso2c" = "iso2c"))
all2 <- inner_join(all1, select(lnd_agri_all, iso2c, AG.LND.AGRI.K2), by = c("iso2c" = "iso2c"))
all3 <- inner_join(all2, select(rur_totl_all, iso2c, SP.RUR.TOTL.ZS), by = c("iso2c" = "iso2c"))

head(all3)
```

```
## X iso2c country SP.POP.TOTL year iso3c region
## 1 1 AR Argentina 44938712 2019 ARG Latin America & Caribbean
## 2 2 BD Bangladesh 163046161 2019 BGD South Asia
## 3 3 BR Brazil 211049527 2019 BRA Latin America & Caribbean
## 4 4 CD Congo, Dem. Rep. 86790567 2019 COD Sub-Saharan Africa
## 5 5 CN China 1397715000 2019 CHN East Asia & Pacific
## 6 6 CO Colombia 50339443 2019 COL Latin America & Caribbean
## capital longitude latitude income lending
## 1 Buenos Aires -58.4173 -34.61180 Upper middle income IBRD
## 2 Dhaka 90.4113 23.70550 Lower middle income IDA
## 3 Brasilia -47.9292 -15.78010 Upper middle income IBRD
## 4 Kinshasa 15.3222 -4.32500 Low income IDA
## 5 Beijing 116.2860 40.04950 Upper middle income IBRD
## 6 Bogota -74.0820 4.60987 Upper middle income IBRD
## NY.GDP.PCAP.PP.CD AG.LND.TOTL.K2 AG.LND.AGRI.K2 SP.RUR.TOTL.ZS
## 1 22947.140 2736690 1487000.0 8.009
## 2 4950.744 130170 91942.1 62.595
## 3 15258.851 8358140 2835460.0 13.176
## 4 1143.453 2267050 262000.0 54.954
## 5 16784.659 9388210 5277330.0 39.692
## 6 15643.731 1109500 446656.0 18.896
```

```
summary(all3)
```

```
## X iso2c country SP.POP.TOTL
## Min. : 1.0 Length:35 Length:35 Min. :4.281e+07
## 1st Qu.: 9.5 Class :character Class :character 1st Qu.:5.331e+07
## Median :18.0 Mode :character Mode :character Median :8.313e+07
## Mean :18.0 Mean :1.760e+08
## 3rd Qu.:26.5 3rd Qu.:1.360e+08
## Max. :35.0 Max. :1.398e+09
##
## year iso3c region capital
## Min. :2019 Length:35 Length:35 Length:35
## 1st Qu.:2019 Class :character Class :character Class :character
## Median :2019 Mode :character Mode :character Mode :character
## Mean :2019
## 3rd Qu.:2019
## Max. :2019
##
## longitude latitude income lending
## Min. : -99.13 Min. : -34.612 Length:35 Length:35
## 1st Qu.: 5.27 1st Qu.: 6.816 Class :character Class :character
## Median : 32.54 Median : 23.706 Mode :character Mode :character
## Mean : 35.29 Mean : 21.536
## 3rd Qu.: 83.82 3rd Qu.: 39.302
## Max. :139.77 Max. : 55.756
##
## NY.GDP.PCAP.PP.CD AG.LND.TOTL.K2 AG.LND.AGRI.K2 SP.RUR.TOTL.ZS
## Min. : 1143 Min. : 97489 Min. : 17008 Min. : 8.009
## 1st Qu.: 5350 1st Qu.: 398311 1st Qu.: 155360 1st Qu.:19.495
## Median :13170 Median : 828340 Median : 368440 Median :33.144
## Mean :20129 Mean : 2126901 Mean : 781561 Mean :40.115
## 3rd Qu.:28855 3rd Qu.: 1910855 3rd Qu.: 694931 3rd Qu.:62.844
## Max. :65118 Max. :16376870 Max. :5277330 Max. :78.775
```

```
## NA's :1      NA's :1
write.csv(all3, file = "all3.csv")
```

Порахуємо як 35 країн з найбільшим населенням розподілилися за регіонами

```
all %>%
  group_by(region) %>%
  count()

## # A tibble: 7 x 2
## # Groups:   region [7]
##   region          n
##   <chr>         <int>
## 1 East Asia & Pacific      8
## 2 Europe & Central Asia    8
## 3 Latin America & Caribbean 4
## 4 Middle East & North Africa 3
## 5 North America           1
## 6 South Asia               3
## 7 Sub-Saharan Africa       8
```

а також за рівнем доходів

```
all %>%
  group_by(income) %>%
  count()

## # A tibble: 4 x 2
## # Groups:   income [4]
##   income          n
##   <chr>         <int>
## 1 High income      8
## 2 Low income       4
## 3 Lower middle income 12
## 4 Upper middle income 11
```

Створимо факторну змінну рівень с.г. населення та подивимось розподіл по рівнях

- низький < 25%
- нижче середнього >=25% - <50%
- вище середнього >=50% - <75%
- високий >=75%

```
all3 %>%
  mutate(rur_level=case_when(SP.RUR.TOTL.ZS<25 ~ "низький",
                             SP.RUR.TOTL.ZS<50 ~ "нижче середнього",
                             SP.RUR.TOTL.ZS<75 ~ "вище середнього",
                             TRUE ~ "високий")) %>%
  group_by(rur_level) %>%
  count()

## # A tibble: 4 x 2
## # Groups:   rur_level [4]
##   rur_level          n
```

```
##   <chr>           <int>
## 1 високий         2
## 2 вище середнього 11
## 3 нижче середнього 9
## 4 низький         13
```

а також які країни потрапили в групу

```
options(dplyr.summarise.inform = FALSE)
```

```
p <- function(v) {
  Reduce(f=paste, x = v)
}
```

```
all3 %>%
  mutate(rur_level=case_when(SP.RUR.TOTL.ZS<25 ~ "низький",
                             SP.RUR.TOTL.ZS<50 ~ "нижче середнього",
                             SP.RUR.TOTL.ZS<75 ~ "вище середнього",
                             TRUE ~ "високий")) %>%
  group_by(rur_level) %>%
  summarize(countries = p(as.character(country)))
```

```
## # A tibble: 4 x 2
##   rur_level countries
##   <chr>      <chr>
## 1 високий   Ethiopia Uganda
## 2 вище середнього Bangladesh Congo, Dem. Rep. Egypt, Arab Rep. India Kenya Myan~
## 3 нижче середньо~ China Algeria Indonesia Italy Nigeria Russian Federation Thai~
## 4 низький   Argentina Brazil Colombia Germany Spain France United Kingdom~
```