

Самостійна робота 3

Юрій Харченко

11/25/2020

Продовження роботи, розпочатої в Самостійна робота 2

Підключимо потрібні пакети

```
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
```

```
library(psych)
```

```
library(doBy)
```

```
##  
## Attaching package: 'doBy'  
## The following object is masked from 'package:dplyr':  
##  
##   order_by
```

```
library(pander)
```

Завантажимо створений в попередній роботі фрейм

```
all <- read.csv(file = "all3.csv")  
  
all %>%  
  select(  
    "N"=X,  
    "Країна"=country,  
    "Населення" = SP.POP.TOTL,  
    "Сільське населення відс" = SP.RUR.TOTL.ZS,  
    "ВВП на душу долл" = NY.GDP.PCAP.PP.CD,  
    "Площа кв.км" = AG.LND.TOTL.K2,  
    "Площа сг кв.км" = AG.LND.AGRI.K2  
  ) %>%  
  pander(style = "grid", use.hyphenning = TRUE, split.cells = 3)
```

N	Країна	Населення	Сільське населення відс	ВВП на душу долл	Площа кв.км	Площа сг кв.км
1	Argentina	44938712	8.009	22947	2736690	1487000
2	Bangladesh	1.63e+08	62.59	4951	130170	91942
3	Brazil	2.11e+08	13.18	15259	8358140	2835460
4	Congo, Dem. Rep.	86790567	54.95	1143	2267050	262000
5	China	1.398e+09	39.69	16785	9388210	5277330
6	Colombia	50339443	18.9	15644	1109500	446656
7	Germany	83132799	22.62	56052	349360	166570
8	Algeria	43053054	26.81	11820	2381740	413602
9	Egypt, Arab Rep.	100388073	57.27	12251	995450	37338
10	Spain	47076781	19.43	42214	499564	262657
11	Ethiopia	112078730	78.78	2312	1e+06	362590
12	France	67059887	19.29	49435	547557	287180
13	United Kingdom	66834405	16.35	48710	241930	173500
14	Indonesia	270625568	44.02	12302	1811570	570000
15	India	1.366e+09	65.53	7034	2973190	1797210
16	Iran, Islamic Rep.	82913906	24.61	NA	1628760	459540
17	Italy	60297396	29.26	44197	294140	127170
18	Japan	126264931	8.302	43236	364560	44710
19	Kenya	52573973	72.49	4509	569140	276300
20	Korea, Rep.	51709098	18.57	43029	97489	17008
21	Myanmar	54045420	69.15	5355	653080	127600
22	Mexico	127575529	19.56	20411	1943950	1062360
23	Nigeria	2.01e+08	48.84	5348	910770	708000
24	Philippines	108116615	52.85	9277	298170	124400
25	Pakistan	216565318	63.09	4885	770880	368440
26	Russian Federation	144373535	25.41	29181	16376870	2177218
27	Sudan	42813238	65.06	4123	NA	681862
28	Thailand	69625582	49.31	19228	510890	221100
29	Turkey	83429615	24.37	27875	769630	383270
30	Tanzania	58005463	65.5	2771	885800	396500
31	Ukraine	44385155	30.53	13341	579290	415150
32	Uganda	44269594	75.64	2272	200520	144150
33	United States	328239523	17.54	65118	9147420	4058625
34	Vietnam	96462106	63.37	8374	310070	121780
35	South Africa	58558270	33.14	12999	1213090	968410

```
all %>%
  select(
    "N"=X,
    "Region"=region,
    "Рівень доходу" = income,
    "Група позик" = lending
  ) %>%
  pander(style = "grid", use.hyphening = TRUE, split.cells = 3)
```

N	Perion	Рівень доходу	Група позик
1	Latin America & Caribbean	Upper middle income	IBRD
2	South Asia	Lower middle income	IDA
3	Latin America & Caribbean	Upper middle income	IBRD
4	Sub-Saharan Africa	Low income	IDA
5	East Asia & Pacific	Upper middle income	IBRD
6	Latin America & Caribbean	Upper middle income	IBRD
7	Europe & Central Asia	High income	Not classified
8	Middle East & North Africa	Upper middle income	IBRD
9	Middle East & North Africa	Lower middle income	IBRD
10	Europe & Central Asia	High income	Not classified
11	Sub-Saharan Africa	Low income	IDA
12	Europe & Central Asia	High income	Not classified
13	Europe & Central Asia	High income	Not classified
14	East Asia & Pacific	Lower middle income	IBRD
15	South Asia	Lower middle income	IBRD
16	Middle East & North Africa	Upper middle income	IBRD
17	Europe & Central Asia	High income	Not classified
18	East Asia & Pacific	High income	Not classified
19	Sub-Saharan Africa	Lower middle income	Blend
20	East Asia & Pacific	High income	Not classified
21	East Asia & Pacific	Lower middle income	IDA

N	Perion	Рівень доходу	Група позик
22	Latin America & Caribbean	Upper middle income	IBRD
23	Sub-Saharan Africa	Lower middle income	Blend
24	East Asia & Pacific	Lower middle income	IBRD
25	South Asia	Lower middle income	Blend
26	Europe & Central Asia	Upper middle income	IBRD
27	Sub-Saharan Africa	Lower middle income	IDA
28	East Asia & Pacific	Upper middle income	IBRD
29	Europe & Central Asia	Upper middle income	IBRD
30	Sub-Saharan Africa	Low income	IDA
31	Europe & Central Asia	Lower middle income	IBRD
32	Sub-Saharan Africa	Low income	IDA
33	North America	High income	Not classified
34	East Asia & Pacific	Lower middle income	IBRD
35	Sub-Saharan Africa	Upper middle income	IBRD

Коментар до завантажених даних

Дані отримані за допомогою пакету WDI, який дозволяє робити он-лайн запити до баз даних Світового банку all3 - містить дані 35 країн з найбільшим населенням з такими кількісними показниками:

- SP.POP.TOTL - загальна кількість населення
- SP.RUR.TOTL.ZS - сільське населення у % до загального
- NY.GDP.PCAP.PP.CD - ВВП на душу населення, \$
- AG.LND.TOTL.K2 - площа, кв.км
- AG.LND.AGRI.K2 - площа сільсько-господарська, кв.км

та факторними змінними:

- region - регіон
- income - рівень доходу

- lending - група позик

Форматування таблиць виконано за допомогою пакету pander

Описові статистики за допомогою пакету psych

describe - описові статистики

```
# describe

all %>%
  select("Населення" = SP.POP.TOTL,
         "Сільське населення відс" = SP.RUR.TOTL.ZS,
         "ВВП" = NY.GDP.PCAP.PP.CD,
         "Площа кв.км" = AG.LND.TOTL.K2,
         "Площа сг кв.км" = AG.LND.AGRI.K2) %>%
  describe(na.rm = TRUE) %>%
  as.data.frame() %>%
  pander(style = "grid", use.hyphening = TRUE, split.cells = 3)
```

Table 3: Table continues below

	vars	n	mean	sd	median	trimmed
Населення	1	35	1.76e+08	3.09e+08	83132799	101352626
Сільське населення відс	2	35	40.12	22.01	33.14	39.57
ВВП	3	34	20129	18037	13170	18145
Площа кв.км	4	34	2126901	3494435	828340	1320499
Площа сг кв.км	5	35	781561	1167531	368440	520143

Table 4: Table continues below

	mad	min	max	range	skew
Населення	48619430	42813238	1.398e+09	1.355e+09	3.368
Сільське населення відс	23.96	8.009	78.78	70.77	0.2153
ВВП	12562	1143	65118	63975	0.9212
Площа кв.км	777209	97489	16376870	16279381	2.598
Площа сг кв.км	357069	17008	5277330	5260322	2.411

	kurtosis	se
Населення	10.28	52238037
Сільське населення відс	-1.499	3.721
ВВП	-0.4595	3093
Площа кв.км	6.491	599291
Площа сг кв.км	5.452	197349

Описові статистики за допомогою пакету doBy

summary_by - вибіркове середнє по регіонах

```
# summaryBy
fun1 <- function(x){
  c(
    mean=mean(x)
  )}
all %>%
  select("Region" = region,
         "Населення" = SP.POP.TOTL
        ) %>%
  summary_by(formula = list("Населення", "Region"),
             FUN = fun1
            ) %>%
  pander(style = "grid", use.hyphenning = TRUE, split.cells = 3)
```

Region	Населення.mean
East Asia & Pacific	271820540
Europe & Central Asia	74573697
Latin America & Caribbean	108475803
Middle East & North Africa	75451678
North America	328239523
South Asia	5.82e+08
Sub-Saharan Africa	82006679

summary_by - скільки країн з вибірки по регіонах

```
# summaryBy
fun2 <- function(x){
  c(
    n=length(x)
  )}
all %>%
  select("Region" = region,
         "N" = SP.POP.TOTL
        ) %>%
  summary_by(formula = list("N", "Region"),
             FUN = fun2,
             keep.names = TRUE
            ) %>%
  pander(style = "grid", use.hyphenning = TRUE, split.cells = 3)
```

Region	N
East Asia & Pacific	8
Europe & Central Asia	8
Latin America & Caribbean	4

Періон	N
Middle East & North Africa	3
North America	1
South Asia	3
Sub-Saharan Africa	8

Найбільша кількість (8) країн з великим числом населення в таких регіонах, як

- East Asia & Pacific
- Europe & Central Asia
- Sub-Saharan Africa