

인구수 데이터 분석

. 세계은행 데이터셋

제출일	2021, 04, 01
작성자	정승호

목 차

서론	3
본론	
1. 2019년 대비 국가 총 인구수와 기타 데이터 분석	
1.1. 2019년 기준 나라별 총 인구수	3
1.2. 그 나라에서 가장 큰 도시에 사는 인구수	5
1.3. 나라별 총 인구수 대비 그 나라에서 가장 큰 도시에 사는 인구의 비율	6
2. 총 인구수 대비 가장 큰 도시에 거주하는 비율이 높은 10개국	7
3. 총 인구수 대비 가장 큰 도시에 거주하는 비율이 낮은 10개국	8
4. 데이터가 누락된 국가들의 리스트	9
결론	10

■ 서론

이 레포트는 세계은행(World Bank)에서 집계한 세계 264개국 인구와 관련된 다양한 통계량이 제공된 데이터셋을 이용하여 R code로 데이터 분석을 수행한 결과를 다룬다.

■ 본문

(1) 2019년 기준 나라별 총 인구수, 그 나라에서 가장 큰 도시에 사는 인구수, 나라별 총 인구수 대비 그 나라에서 가장 큰 도시에 사는 인구의 비율 분석

(1)-1 2019년 기준 국가별 총 인구수

소스코드

```
#####
#1-1 264국 2019년 기준 총 인구수 구하기
#####
COUNTRY <- c()
TOTAL_POPULATION = c()
for(i in 1:265){
  dataSource = worldbankData %>% filter(Country.Code == COUNTRYCODE[i])
  populationData = dataSource %>% filter(Series.Name%in%c("Population ages 0-14, total",
    "Population ages 15-64, total",
    "Population ages 65 and above, total"))
  NA.CHECKER = which(populationData$X2019..YR2019=="..") #조건식의 결과가 참이면 정수형이 반환되고, 거짓이면 0
  #이 반환=".."면 NA임
  if(sum(NA.CHECKER)==6){TOTAL = NA}else{TOTAL = sum(as.numeric(populationData$X2019..YR2019))}#NA체크값
  #이 6 = 인구수 X란뜻
  COUNTRY = c(COUNTRY,COUNTRYNAME[i])
  TOTAL_POPULATION = c(TOTAL_POPULATION,TOTAL)}
popResult <- data.frame(COUNTRY,TOTAL_POPULATION)
popResult
```

결과값

▲	COUNTRY	TOTAL_POPULATION
1	Afghanistan	38041754
2	Albania	2854191
3	Algeria	43053054
4	American Samoa	NA
5	Andorra	NA
6	Angola	31825295
7	Antigua and Barbuda	97118
8	Argentina	44938712
9	Armenia	2957731
10	Aruba	106314
11	Australia	25364306
12	Austria	8877067
13	Azerbaijan	10023318
14	Bahamas, The	389482
15	Bahrain	1641172
16	Bangladesh	163046161
17	Barbados	287025
18	Belarus	9466856
19	Belgium	11484055

• • •

▲	COUNTRY	TOTAL_POPULATION
248	Middle East & North Africa (IDA & IBRD countries)	384771780
249	Middle income	5767234366
250	North America	365828785
251	Not classified	NA
252	OECD members	1359963501
253	Other small states	31326547
254	Pacific island small states	2392670
255	Post-demographic dividend	1112914459
256	Pre-demographic dividend	944902749
257	Small states	40995958
258	South Asia	1835776742
259	South Asia (IDA & IBRD)	1835776742
260	Sub-Saharan Africa	1106957898
261	Sub-Saharan Africa (excluding high income)	1105594562
262	Sub-Saharan Africa (IDA & IBRD countries)	1106957898
263	Upper middle income	2853870975
264	World	7670714537
265		0

(1)-2 각 국가별 가장 큰 도시에 사는 인구수

소스코드

```
#####
#1-2 2019년 기준 그 나라에서 가장 큰 도시에 사는 인구수
#####
worldbankData <- read.csv("worldbank1.csv",header = T, na.strings=c("..", "NA",""))
#불러올 때 결측값 NA로 표기

#Population in largest city의 2019년 기준 데이터 추출
worldbankData %>%
  filter(Series.Name == "Population in largest city") %>%
  select(Country.Code, Series.Name, X2019..YR2019.)
```

결과값

	Country.Code	Series.Name	X2019..YR2019.
1	AFG	Population in largest city	4114030
2	ALB	Population in largest city	484624
3	DZA	Population in largest city	2729325
4	ASM	Population in largest city	NA
5	AND	Population in largest city	NA
6	AGO	Population in largest city	8044735
7	ATG	Population in largest city	NA
8	ARG	Population in largest city	15057273
9	ARM	Population in largest city	1083323
10	ABW	Population in largest city	NA
11	AUS	Population in largest city	4870388
12	AUT	Population in largest city	1915338
13	AZE	Population in largest city	2313138
14	BHS	Population in largest city	NA
15	BHR	Population in largest city	600433
16	BGD	Population in largest city	20283552
17	BRB	Population in largest city	NA
18	BLR	Population in largest city	2016677
19	BEL	Population in largest city	2065284

• • •

	Country.Code	Series.Name	X2019..YR2019.
247	MNA	Population in largest city	NA
248	TMN	Population in largest city	NA
249	MIC	Population in largest city	NA
250	NAC	Population in largest city	NA
251	INX	Population in largest city	NA
252	OED	Population in largest city	NA
253	OSS	Population in largest city	NA
254	PSS	Population in largest city	NA
255	PST	Population in largest city	NA
256	PRE	Population in largest city	NA
257	SST	Population in largest city	NA
258	SAS	Population in largest city	NA
259	TSA	Population in largest city	NA
260	SSF	Population in largest city	NA
261	SSA	Population in largest city	NA
262	TSS	Population in largest city	NA
263	UMC	Population in largest city	NA
264	WLD	Population in largest city	NA

(1)-3 국가별 총 인구수 대비 가장 큰 도시에 사는 인구의 비율

소스코드

```
#####
#1-3 나라별 총 인구수 대비 그 나라에서 가장 큰 도시에 사는 인구의 비율
#####
COUNTRY <- c()
CITY_RATIO <- c()
for(i in 1:265){
  dataSource = worldbankData %>% filter(Country.Code == COUNTRYCODE[i])
  ratioLcity = dataSource %>% filter(Series.Name == "Population in the largest city (% of urban
population)")
  NA.CHECKER = which(ratioLcity$X2019..YR2019.=="..") #조건식의 결과가 참이면 정수형이 반환되고, 거짓이면 0이
반환=".."면 NA임
  if(length(NA.CHECKER)==1){RATIO = NA}else{RATIO = ratioLcity$X2019..YR2019.}#NA체크값이 1 = 데이터가 NA
  COUNTRY = c(COUNTRY,COUNTRYNAME[i])
  CITY_RATIO = c(CITY_RATIO,RATIO)
}
COUNTRY <- COUNTRY[-265]
ratioResult <- data.frame(COUNTRY,CITY_RATIO)
ratioResult
```

결과값

	COUNTRY	CITY_RATIO
1	Afghanistan	41.991583
2	Albania	27.730942
3	Algeria	8.661747
4	American Samoa	NA
5	Andorra	NA
6	Angola	38.197262
7	Antigua and Barbuda	NA
8	Argentina	36.423390
9	Armenia	57.936421
10	Aruba	NA
11	Australia	22.294193
12	Austria	36.861175
13	Azerbaijan	41.183177
14	Bahamas, The	NA
15	Bahrain	40.926271
16	Bangladesh	33.258585
17	Barbados	NA
18	Belarus	27.090417
19	Belgium	18.313532

• • •

	COUNTRY	CITY_RATIO
247	Middle East & North Africa (excluding high income)	23.814102
248	Middle East & North Africa (IDA & IBRD countries)	23.881286
249	Middle income	13.900253
250	North America	8.278975
251	Not classified	NA
252	OECD members	18.002049
253	Other small states	NA
254	Pacific island small states	NA
255	Post-demographic dividend	17.470630
256	Pre-demographic dividend	28.405001
257	Small states	NA
258	South Asia	11.324374
259	South Asia (IDA & IBRD)	11.324374
260	Sub-Saharan Africa	26.434384
261	Sub-Saharan Africa (excluding high income)	26.434384
262	Sub-Saharan Africa (IDA & IBRD countries)	26.434384
263	Upper middle income	11.855436
264	World	16.099026

(2) 총 인구수 대비 각 국가별 가장 큰 도시에 사는 인구의 비율이 가장 높은 10개국의 리스트 작성.

소스코드

```
#####
#2. 총 인구수 대비 그 나라에서 가장 큰 도시에 사는 인구의 비율이 가장 높은 나라 10개국
#####
COUNTRY <- c() #나라
CITY_RATIO <- c() #비율
for(i in 1:265){
  dataSource = worldbankData %>% filter(Country.Code == COUNTRYCODE[i])
  ratioLcity = dataSource %>% filter(Series.Name == "Population in the largest city (% of urban
population)")
  NA.CHECKER = which(ratioLcity$X2019..YR2019=="..") #조건식의 결과가 참이면 정수형이 반환되고, 거짓이면 0이
반환=.."면 NA임
  if(length(NA.CHECKER)==1){RATIO = NA}else{RATIO = ratioLcity$X2019..YR2019.}#NA체크값이 1 = 데이터가 NA
  COUNTRY = c(COUNTRY,COUNTRYNAME[i])
  CITY_RATIO = c(CITY_RATIO,RATIO)
}
COUNTRY <- COUNTRY[-265]
ratioResult <- data.frame(COUNTRY,CITY_RATIO)
ratioResult1 = na.omit(ratioResult) %>% arrange(desc(as.numeric(CITY_RATIO))) %>% slice(1:10)
ratioResult1
```

결과값

	COUNTRY	CITY_RATIO
1	Macao SAR, China	100.00000
2	Singapore	100.00000
3	Hong Kong SAR, China	99.77857
4	Puerto Rico	82.02609
5	Paraguay	75.22476
6	Djibouti	74.98580
7	Trinidad and Tobago	73.35581
8	Kuwait	72.55604
9	Mongolia	70.23594
10	Guinea-Bissau	68.83266

(3) 총 인구수 대비 각 국가별 가장 큰 도시에 사는 인구의 비율이 가장 낮은 10개국의 리스트를 작성.

소스코드

```
#####
#3 총 인구수 대비 그 나라에서 가장 큰 도시에 사는 비율이 가장 낮은 나라 10개국
#####
COUNTRY <- c()
CITY_RATIO <- c()
for(i in 1:265){
  dataSource = worldbankData %>% filter(Country.Code == COUNTRYCODE[i])
  ratioLcity = dataSource %>% filter(Series.Name == "Population in the largest city (% of urban
population)")
  NA.CHECKER = which(ratioLcity$X2019..YR2019=="..") #조건식의 결과가 참이면 정수형이 반환되고, 거짓이면 0이
반환=".."면 NA임
  if(length(NA.CHECKER)==1){RATIO = NA}else{RATIO = ratioLcity$X2019..YR2019.}#NA체크값이 1 = 데이터가 NA
  COUNTRY = c(COUNTRY,COUNTRYNAME[i])
  CITY_RATIO = c(CITY_RATIO,RATIO)
}
COUNTRY <- COUNTRY[-265]
ratioResult <- data.frame(COUNTRY,CITY_RATIO)
ratioResult
ratioResult2 = na.omit(ratioResult) %>% arrange(as.numeric(CITY_RATIO)) %>% slice(1:10)
ratioResult2
str(ratioResult)
```

결과값

	COUNTRY	CITY_RATIO
1	China	3.122084
2	Germany	5.532074
3	India	6.241438
4	United States	6.947651
5	Indonesia	7.021786
6	Netherlands	7.155842
7	East Asia & Pacific (IDA & IBRD countries)	7.455967
8	East Asia & Pacific (excluding high income)	7.614195
9	Poland	7.791458
10	North America	8.278975

(4) 데이터가 누락된 국가의 리스트를 작성.

소스코드

```
#####  
#4번 해당 항목에 데이터가 없는 나라의 리스트를 작성하시오  
#####  
Missingvalue <- table(is.na(ratioResult$CITY_RATIO)) #결측값 개수 확인  
Missingvalue  
  
ratioResult2 <- ratioResult %>% filter(is.na(CITY_RATIO)) #결측값이 있는 나라 추출  
View(ratioResult2)
```

결과값

	COUNTRY	CITY_RATIO			
			20	Eritrea	NA
1	American Samoa	NA	21	Eswatini	NA
2	Andorra	NA	22	Faroe Islands	NA
3	Antigua and Barbuda	NA	23	Fiji	NA
4	Aruba	NA	24	French Polynesia	NA
5	Bahamas, The	NA	25	Gibraltar	NA
6	Barbados	NA	26	Greenland	NA
7	Belize	NA	27	Grenada	NA
8	Bermuda	NA	28	Guam	NA
9	Bhutan	NA	29	Guyana	NA
10	Botswana	NA	30	Iceland	NA
11	British Virgin Islands	NA	31	Isle of Man	NA
12	Brunei Darussalam	NA	32	Kiribati	NA
13	Cabo Verde	NA	33	Kosovo	NA
14	Cayman Islands	NA	34	Lesotho	NA
15	Channel Islands	NA	35	Liechtenstein	NA
16	Comoros	NA	36	Luxembourg	NA
17	Curacao	NA	37	Maldives	NA
18	Cyprus	NA	38	Malta	NA
19	Dominica	NA			

	COUNTRY	CITY_RATIO		COUNTRY	CITY_RATIO
39	Marshall Islands	NA	53	Slovenia	NA
40	Mauritius	NA	54	Solomon Islands	NA
41	Micronesia, Fed. Sts.	NA	55	St. Kitts and Nevis	NA
42	Monaco	NA	56	St. Lucia	NA
43	Montenegro	NA	57	St. Martin (French part)	NA
44	Nauru	NA	58	St. Vincent and the Grenadines	NA
45	New Caledonia	NA	59	Suriname	NA
46	Northern Mariana Islands	NA	60	Timor-Leste	NA
47	Palau	NA	61	Tonga	NA
48	Samoa	NA	62	Turks and Caicos Islands	NA
49	San Marino	NA	63	Tuvalu	NA
50	Sao Tome and Principe	NA	64	Vanuatu	NA
51	Seychelles	NA	65	Virgin Islands (U.S.)	NA
52	Sint Maarten (Dutch part)	NA	66	Caribbean small states	NA
53	Slovenia	NA	67	Not classified	NA
54	Solomon Islands	NA	68	Other small states	NA
55	St. Kitts and Nevis	NA	69	Pacific island small states	NA
56	St. Lucia	NA	70	Small states	NA
57	St. Martin (French part)	NA			

■ 결론

이 레포트를 통해 세계은행(World Bank)에서 집계한 세계 264개국 인구와 관련된 다양한 통계량이 제공된 데이터셋을 R code로 데이터 핸들링을 할 수 있었다.