LearningSpoonsR\_lifeCountry

Geon Seung Lee

2019-01-14

## 1. 자료 불러오기

### 1-1. 폴더 만들기

C:/LS-DS/classProject라는 폴더를 만드세요.

### 1-2. Rmd 파일 만들기

Rstudio를 열어서 File -> New File -> R Script를 하면 새로운 소스 파일이 생성됩니다.  
RStudio를 완전히 닫고 탐색기에서 classProject1.Rmd을 더블클릭하여 엽니다.

### 1-3. 데이터셋 받기

[github](https://github.com/LearningSpoonsR/LS-DS/tree/master/classProject)에서 lifeCountry.csv 파일을 다운받아서 위 폴더에 넣으세요.

이제 분석을 위한 모든 준비가 끝났습니다.

### 1-4. 데이터셋 R로 불러오기

실습데이터를 dataset으로 불러오도록 하겠습니다.

dataset <- read.csv("lifeCountry.csv", stringsAsFactors = FALSE)  
library(dplyr)  
library(ggplot2)

## 2. 실습하기

### 2-1. 데이터는 총 몇개의 행과 열로 되어 있습니까?

데이터 구조 확인

str(dataset)

## 'data.frame': 161 obs. of 11 variables:  
## $ X : int 1 2 3 4 5 6 7 8 9 10 ...  
## $ Country : chr "Afghanistan" "Albania" "Algeria" "Angola" ...  
## $ ISO3 : chr "AFG" "ALB" "DZA" "AGO" ...  
## $ Continent : chr "Asia" "Europe" "Africa" "Africa" ...  
## $ OECD\_GDP : num NA NA NA NA NA ...  
## $ GDP\_per\_Capita : int 572 4520 4225 4401 16826 14061 3690 3690 56135 46435 ...  
## $ Population : int 35530081 2930187 41318142 29784193 102012 44271041 2930450 2930450 24450561 8735453 ...  
## $ Life.Expectancy: num 60.5 77.8 75.6 52.4 76.4 76.3 74.8 74.8 82.8 81.5 ...  
## $ Female : num 61.9 80.7 77.5 54 78.6 79.9 77.7 77.7 84.8 83.9 ...  
## $ Male : num 59.3 75.1 73.8 50.9 74.1 72.7 71.6 71.6 80.9 79 ...  
## $ Health.Adjusted: num 52.2 68.8 66 45.8 67.5 67.6 66.8 66.8 71.9 72 ...

데이터는 총 161개의 행과 11개의 열로 이루어져 있습니다.

### 2-2. 가장 GDP가 높고 낮은 나라는 어디인가요? 가장 기대 수명이 길고 짧은 나라는 어디인가요?

가장 GDP가 높은 나라는 Luxembourg 입니다.  
**Luxembourg**의 Data는 아래와 같습니다.

library(pander)  
library(knitr)  
library(kableExtra)  
dataset %>% arrange(desc(GDP\_per\_Capita)) %>% head(1) %>% pander()

Table continues below

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| X | Country | ISO3 | Continent | OECD\_GDP | GDP\_per\_Capita | Population |
| 90 | Luxembourg | LUX | Europe | 89067 | 107708 | 583455 |

|  |  |  |  |
| --- | --- | --- | --- |
| Life.Expectancy | Female | Male | Health.Adjusted |
| 82 | 84 | 79.8 | 71.8 |

# %>% kable() # %>% kable\_styling(bootstrap\_options = "striped", full\_width = F, position = "left")

가장 GDP가 낮은 나라는 South Sudan 입니다.  
**South Sudan**의 Data는 아래와 같습니다.

dataset %>% arrange(GDP\_per\_Capita) %>% head(1) %>% kable() %>% kable\_styling(bootstrap\_options = "striped", full\_width = F, position = "left")

가장 기대수명이 높은 나라는 Japan 입니다.  
**Japan**의 Data는 아래와 같습니다.

dataset %>% arrange(desc(Life.Expectancy)) %>% head(1) %>% kable() %>% kable\_styling(bootstrap\_options = "striped", full\_width = F, position = "left")

가장 기대수명이 낮은 나라는 Sierra Leone 입니다.  
**Sierra Leone**의 Data는 아래와 같습니다.

dataset %>% arrange(Life.Expectancy) %>% head(1) %>% kable() %>% kable\_styling(bootstrap\_options = "striped", full\_width = F, position = "left")

### 2-3. 대륙별로 GDP와 기대수명의 평균을 구해보세요.

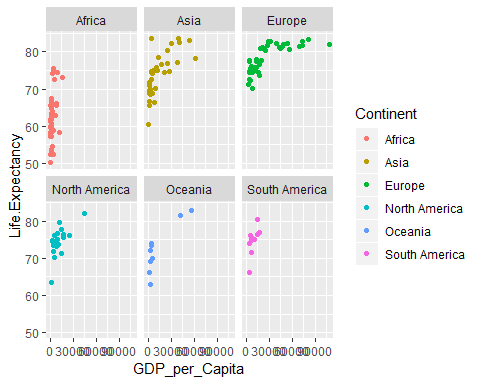
dataset %>% group\_by(Continent) %>% summarise(avg\_GDP = mean(GDP\_per\_Capita), avg\_LE = mean(Life.Expectancy)) %>% arrange(desc(avg\_GDP)) %>% kable() %>% kable\_styling(bootstrap\_options = "striped", full\_width = F, position = "left")

### 2-4. 각 나라의 GDP를 X축으로, 기대수명을 Y축으로 산점도를 그리고 대륙에 따라서 점의 색깔이 달라지게 해보세요.

ggplot(data = dataset, mapping = aes(x=GDP\_per\_Capita, y=Life.Expectancy))+  
 geom\_point(aes(color=Continent))

### 2-5. facet을 이용해서 대륙별로 GDP와 기대수명에 대한 산점도를 그려보세요.

a <- ggplot(data = dataset, mapping = aes(x=GDP\_per\_Capita, y=Life.Expectancy))+  
 geom\_point(aes(color=Continent))  
a <- a+facet\_wrap(~ Continent)  
print(a)



### 2-6. 국가별로 성별에 따라 기대수명이 다릅니다. mutate 함수를 사용해서 ageSexDiff라는 변수를 만들어보세요.

dataset<-dataset %>% mutate(ageSexDiff = abs(Female-Male))  
dataset %>% arrange(desc(ageSexDiff)) %>% head()

## X Country ISO3 Continent OECD\_GDP GDP\_per\_Capita Population  
## 1 17 Belarus BLR Europe NA 5585 9468338  
## 2 89 Lithuania LTU Europe 28116.83 16443 2890297  
## 3 124 Rwanda RWA Africa NA 754 12208407  
## 4 158 Ukraine UKR Europe NA 2458 44222947  
## 5 84 Latvia LVA Europe 24243.70 15402 1949670  
## 6 49 Estonia EST Europe 27929.45 19618 1309632  
## Life.Expectancy Female Male Health.Adjusted ageSexDiff  
## 1 72.3 78.0 66.5 65.1 11.5  
## 2 73.6 79.1 68.1 66.0 11.0  
## 3 66.1 71.1 60.9 56.6 10.2  
## 4 71.3 76.1 66.3 64.1 9.8  
## 5 74.6 79.2 69.6 67.1 9.6  
## 6 77.6 82.0 72.7 68.9 9.3

### 2-7. 어떤 대륙에서 ageSexDiff가 가장 크고 작은가요?

dataset %>% group\_by(Continent) %>% summarise(avg\_ageSexDiff=mean(ageSexDiff)) %>% arrange(desc(avg\_ageSexDiff))

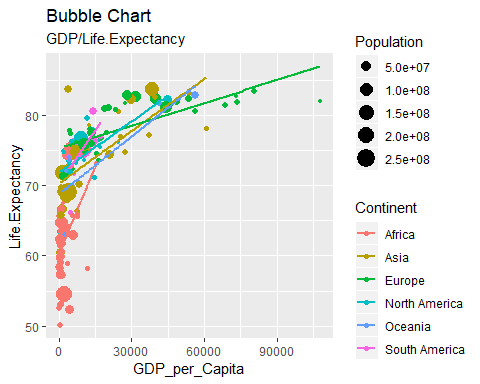
## # A tibble: 6 x 2  
## Continent avg\_ageSexDiff  
## <chr> <dbl>  
## 1 Europe 6.00  
## 2 South America 5.97  
## 3 North America 5.52  
## 4 Asia 4.86  
## 5 Oceania 4.71  
## 6 Africa 3.72

## 3. 자유로운 분석

### 3-1. 인구를 포함하여 X축에 GDP, Y축에 기대수명으로 버블차트를 만들어보세요.

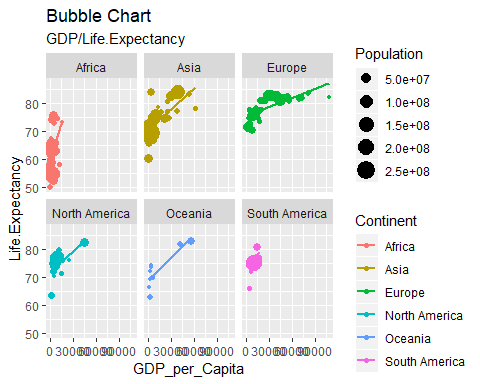
“각 국가 인구별 GDP/기대수명”

b<- ggplot(dataset, mapping = aes(x=GDP\_per\_Capita, y=Life.Expectancy))+  
 labs(subtitle = "GDP/Life.Expectancy",  
 title= "Bubble Chart")  
b + geom\_jitter(aes(col=Continent, size=Population)) +  
 geom\_smooth(aes(col=Continent), method="lm", se=F)



### 3-2. 각 대륙별로 버블차트를 만들어보세요.

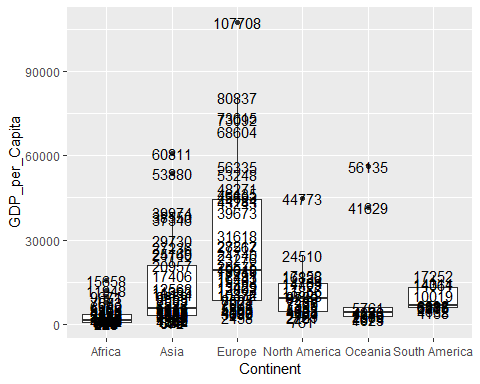
b + geom\_jitter(aes(col=Continent, size=Population)) +  
 geom\_smooth(aes(col=Continent), method="lm", se=F) +  
 facet\_wrap( ~ Continent)



### 3-3. 각 대륙별 GDP 및 기대수명에 대한 박스플롯을 만들어보세요.

1. GDP BOX PLOT

ggplot(dataset, aes(x=Continent, y=GDP\_per\_Capita)) +  
 geom\_boxplot() +  
 geom\_text(aes(label=GDP\_per\_Capita))



# 박스플롯에서 아웃라이러를 제외한 최대/상위25/중간값/하위25/최소 데이터값만 표기하고 싶습니다.